THE ANNALS *of* MATHEMATICAL STATISTICS

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Volume VIII, Number 2
June, 1937

PUBLISHED QUARTERLY
ANN ARBOR, MICHIGAN

Four Dollars per annum

Back numbers available at the following prices:
Vols. I–IV \$5 each. Single numbers \$1.50
Vol. V to date \$4 each. Single numbers \$1.25

Made in United States of America

Address: ANNALS OF MATHEMATICAL STATISTICS
Post Office Box 171, Ann Arbor, Michigan

Office of the Institute of Mathematical Statistics:
Secretary: ALLEN T. CRAIG, University of Iowa
Iowa City, Iowa

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, INC.
BALTIMORE, MD.

REGRESSION AND CORRELATION EVALUATED BY A METHOD OF PARTIAL SUMS

By Felix Bernstein

"To be sure, Laplace viewed the matter in a similar way but he selected the absolute value of the error as a measure of loss. But if we mistake not, this position is certainly not less arbitrary than our own; that is to say, whether the double error is to be considered just as tolerable as, or worse than, the simple error twice repeated and whether it is thus more fitting to ascribe to the double error only a double weight, or a greater one, is a question which is neither in itself clear nor determinable by mathematical proof but has to be left entirely to individual discretion.

"Furthermore, it cannot be denied that the assumption under discussion violates the principle of continuity and precisely for this reason the procedure based on it strongly defies analytic treatment while the results to which our principle leads have the advantage of simplicity as well as of generality."—*F. G. Gauss: Theoria combinationis observationum, pars prior, art. 6.*

Since the "Theoria Combinationis" of C. F. Gauss appeared in the year 1821 a century of Mathematical Statistics has been dominated by the ideas of this classical treatise—ideas whose fertility does not seem to be exhausted even today.

The germ of most modern contributions to mathematical statistics—in fact also those of Karl Pearson and his school—go back decidedly to this paper. Though the immediate achievements of Gauss are so conspicuous as not to need any comment, a true critical appreciation of the work can be gained only by comparing it with the previous methods of Laplace, superseded by those of Gauss.

For such critical appreciation, C. F. Gauss himself has prepared the ground in the lines quoted at the beginning of this article. To Gauss the standard deviation is a measure of uncertainty or risk of a game in which the errors of observation are considered as causing only losses. In this he follows the lead of his great predecessor. The difference between them is that Gauss adopts the square of the error as a measure of the loss while Laplace adopts its absolute value for this purpose. Either choice frees the error from its sign so that the loss is the same regardless of the sign of the error.

Gauss considers this choice of the measure of the loss as purely conventional. Therefore he feels justified in adopting the square of the error because in adopting the square instead of the absolute value of the error, the mathematics he uses remains in the easily accessible domain of analytical processes. This creates for these methods a superiority in elegance, simplicity, and generality.

The modern developments of mathematical statistics, based on the principles

of Gauss, have confirmed the correctness of this viewpoint. This has proved true particularly in the theory of analysis of variance developed by R. A. Fisher and in the more general theory of semi-invariants, first defined by N. H. Thiele.

The inadequacy of the Gaussian method seriously impairing its value for statistical use has come to light through the investigations of Karl Pearson of distributions of one and two variables. Since the moments of higher order involve standard deviations of increasing magnitude the characterization of the distributions by means of the moments, in line with the Gauss-Thiele concepts, becomes practically impossible. Therefore it was of the greatest interest that Lindeberg was able to derive an expression for the standard deviation of a measure of skewness constructed not on Gaussian but on Laplacian lines, namely based exclusively upon the sign of the error. The mathematical difficulties surmounted by Lindeberg by a very involved and difficult analysis—with some clearly indicated gaps in the proofs—are precisely of the character of those that Gauss wished to avoid. Encouraged by the success of Lindeberg, I have developed in two papers[1] the standard deviations of more general moments and the correlations between them of which the mean deviation of Laplace and Lindeberg's measure of skewness are special cases. The proofs have been arrived at by a rather simple and rigorous procedure. These new moments, together with the old ones, form a new system of statistical characteristics by which a distribution in one or two variables can be described by expressions of lower order and therefore of greater precision. This method makes unnecessary the use of moments of higher order than the third.

But another point of interest is still involved. It has been assumed that the Gaussian characteristics give a greater amount of information than those of Laplace. This is proved, however, only for the case of the normal distribution $\frac{h}{\sqrt{\pi}} e^{-h^2x^2}$. This was recognized by Gauss himself in his paper of April, 1816, that appeared five years earlier than the Theoria Combinationis Observationum. In article 6 of his paper, he says, that the constant h of a normal distribution obtained from one hundred observations by the use of the standard error is as exact as that obtained from one hundred fourteen observations in which the mean deviation is used. Hence with a given number of observations only the equivalent of 88% of the total are used by the second method. This does not hold true for all distributions. The following theorem can easily be proved: The amount of information as defined above, furnished by the use of the mean deviation is greater, equal to, or less than that furnished by the standard deviation, depending respectively upon whether

[1] Felix Bernstein: "Die mittleren Fehlerquadrate und Korrelationen der Potenzmomente und ihre Anwendung auf Funktionen der Potenzmomente," Metron, Vol. X, N. 3 (Nov. 1932).

Felix Bernstein: "Uber den mittleren Fehler der Potenzmomente." Zeitschr. f. d. ges. Vers.-Wissenschaft, Bard 30, Heft 3, March 1930.

$$(\beta_2 - 1) \gtreqless 4(\beta_0 - 1)$$

where

$$\beta_0 = \frac{\mu_2}{\vartheta^2}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

μ_k the k-th moment and ϑ = the mean deviation.

For example, in the distribution $\frac{h}{2}e^{-h|x|}$, the mean deviaition furnishes a greater amount of information than the standard deviation.[2]

In the present paper, we shall discuss the practical use of expressions for correlation and regression in which the new type of statistics formed along Laplacian lines will be used. These new expressions are of a linear form and can be computed therefore more easily than those of Karl Pearson. The amount of information given by these expressions is less than that given by the expressions of Pearson if the normal law, in two variables, is fulfilled. For other distributions, however, this is not generally true. The determination of the standard deviations of these new expressions is given in Metron.[3]

The application of the new expressions of regression and correlation to grouped data is set forth here for the first time. The method is strongly recommended for all cases in which the data lose reliability with increasing deviations from the mean. Deviations in the new method enter the expressions only in the first degree and not in the second as in the case of Pearson's. It is obvious that the influence of the doubtful extreme readings is, therefore, considerably lessened. Since our expressions are linear, no adjustments for grouping (Sheppard's corrections) are necessary.

It ought to be mentioned here that linear expressions for the measurement of correlation have been set up before.

K. Pearson (Biometrika) and Egon Pearson (Biometrika) have derived an expression called "linear correlation ratio" which in case of linear regression is identical with the correlation coefficient.

K. Pearson also discusses the linear correlation coefficient

$$r = \frac{1}{2}\left(S\frac{ysgx}{xsgx} + S\frac{xsgy}{ysgy}\right),$$

---

[2] To this second type of distribution curves also belongs $y = \psi(x)$ where $x(x)$ is the mean of two Gaussian curves with the same origin, i.e. $\psi(x) = \frac{1}{2}\left(\frac{h}{\sqrt{\pi}}e^{-h^2x^2} + \frac{kh}{\sqrt{\pi}}e^{-k^2h^2x^2}\right)$ $1.6 < k < 3.4$.

I owe this remark and some other valuable suggestions regarding the subject of this paper to Mr. Myron Fuchs.

[3] *Op. cit.*

suggested by Lenz and various other linear expressions, all similar to our expression (1). He finds that they are all equal to his quadratic correlation coefficient in the case of a Gaussian distribution.

However, their expressions were not recommended by those authors for the determination of correlation between quantitative variables, because—

1. No easy and practicable methods were given for their evaluation in the case of grouped data.
2. Their standard deviations were not determined.

We now proceed to define the new formulas and to describe the methods for their evaluation. The proofs are furnished in the Appendix to this paper.

Let r_1 and r_2 denote the regression coefficients of x on y and y on x respectively, and r, as usual, the coefficient of correlation, and by $\bar{x}$ and $\bar{y}$ the arithmetic means of the x's and y's. Let us take $\bar{x}$, $\bar{y}$ as the origin, so that x, y are the deviations from the mean. We have

$$(1)\qquad \begin{gathered} r_1 = \frac{\underset{+y}{S}x}{\underset{+y}{S}y} \quad \text{or} \quad r_1 = \frac{\underset{-y}{S}x}{\underset{-y}{S}y} \\ r_2 \quad \frac{\underset{+x}{S}y}{\underset{+x}{S}x} \quad \text{or} \quad r_2 \quad \frac{\underset{-x}{S}y}{\underset{-x}{S}x} \\ r = \sqrt{r_1 \times r_2} \end{gathered}$$

$\underset{+y}{S}x$ denotes a partial sum of the x's, this sum being extended over all the x's of the observations whose y is positive and the other sums have a corresponding meaning.

It should be noted though that if data occur whose y-deviation is 0 (practically never in a grouped table) one-half of the sum of these x's should be added to $\underset{+y}{S}x$. In the $\underset{+x}{S}$ a similar addition should be made in case observations occur in which x is zero. (See Table IV.)

The formulas (1) and all following ones will be proved in the appendix to this article.[4]

---

[4] Using r_1 and r_2 of (1) the regression lines are $y = r_2x$ and $x = r_1y$. They are those straight lines which fit the data best according to the method of least squares, if the weight of the deviations is taken inversely proportional to the absolute value of the variable. Taking x for instance as the independent variable, r_2 is the value of m which minimizes $S \frac{1}{|x|} (y - mx)^2$ (the sum extended over all data x y).

The standard deviations of r_1 and r_2 are

$$\sigma^2_{r_1} = \frac{r_1^{2r}}{2N}(1 + m(m - 2r)) \qquad \text{where } m = \frac{\underset{+x}{Sx}}{\underset{+y}{Sx}}$$

(2)

$$\sigma^2_{r_2} = \frac{r_2^{2r}}{2N}(1 + n(n - 2r)) \qquad \text{where } n = \frac{\underset{+y}{Sy}}{\underset{+x}{Sy}}$$

We are now going to illustrate the computation of r and for this purpose we shall use a table of Pearson's which gives the correlation between the heights of fathers and daughters.

The totals at the right and lower end of the table are first computed and the bracketed numbers are the sums of the numbers that precede. The means are

$$\bar{x} = \frac{1659.5 - 1179}{1376} = +\frac{480.5}{1376}$$

and

$$\bar{y} = \frac{1650.9 - 1390}{1376} = +\frac{260.5}{1376}$$

whose signs determine on which side of the working mean to "quarter" the table. This quartering is done in Table 1 by the lines *vv* and *hh*. Then the totals above the heavy horizontal separating line *hh* and those to the left of the vertical separating line *vv* are found, e.g. 2, 4.5, 7.25, $\cdots$ and .5, .5, 0, $\cdots$. Multiplying these totals by the respective class marks, we find the outside lines: 18, 36, 50.75, $\cdots$ and 5.5, 5, 0, $\cdots$.

$\underset{-y}{Sx}$ is now $= 1107.5 - 420.5 = 687$, and an adjustment for the fact that a working mean has been used has yet to be made. This adjustment is $\bar{x}N_{-y}$ where N_{-y} is the number of negative y's. ($N_{-y} = 728$.)

We have therefore for the adjusted values

$$\underset{-y}{Sx} = 1107.5 - 420.5 + \frac{260.5}{1376}\cdot 728 = 825.07$$

$$\underset{-y}{Sy} = 1179 + \frac{480.5}{1376}\cdot 728 = 1433.21$$

$$r_1 = .5757 \qquad\qquad r_2 = .5170$$

$$r = .546$$

The standard deviations, according to the formulas (2) are

$$\sigma_{r_1} = .031 \qquad \sigma_{r_2} = .027$$

TABLE 1

Correlation between Heights of Fathers and Daughters

x → Height of Fathers y ↓ Height of Daughters

In Inches

			18	36	50.75	66	206.25	183	202.5	212.5	132.5	(1107.5)	86.25	116.5	115.5	69	23.75	7.5	1.75		(420.25)	
		Totals above line	2	4.5	7.25	11	41.25	45.75	67.5	106.25	132.5		86.25	58.25	38.5	17.25	4.75	1.25	.25		Totals	
	Totals left of line		−9	−8	−7	−6	−5	−4	−3	−2	−1	0	1	2	3	4	5	6	7	8	(728)	
5.5	.5	−11					.25	.25													.5	5.5
5	.5	−10					.25	.25													.5	5
		−9																				
8	1	−8								1											1	8
31.5	4.5	−7	.25	.25		.25	1.25	.5		1	.5	.5									4.5	31.5
81	13.5	−6	.25	.25	.5	1.5	4.5	1	1.5	1.5	2.5		.5	.5							14.5	88
73.75	14.75	−5	.25	.75	.5	.75	.75	1	1.75	1.25	5	2.75	.5	.25							15.5	77.5
163	40.75	−4	.5	1	2		6	4.75	5	6.25	11.75	3.5	3.5	2	1.75	.5					48.5	194
240.75	80.25	−3	.75	.75		2.5	8	6.25	12.5	18.25	20.25	11	9	4.75	2.5	1.25	1.25				99	297
212.5	106.25	−2		.5	1.75	2	9.75	11.5	13	23.75	23.75	20.25	16.5	10.25	4.25	3	1.25				141.5	283
131.75	131.75	−1		1	2.25	2	4.5	12	22.75	26	33	28.25	24.75	14.25	13.75	4.75	.75	.5			190.5	190.5
(952.75)		0			.25	2	6	8.25	11	22.75	35.75	37.25	31.5	26.25	16.25	7.75	1.5	.75	.25		212	(1179)
87.25	87.25	1			.25	2.5	1.75	3.25	9.25	23	18.75	28.5	33	34.25	24.5	11.75	5.5	1	.25	1	198.5	198.5
108.5	54.25	2				.5	1	.5	11	12.25	9.25	19.75	30	26.5	22.25	15	4.75	3.75	2	1	159.5	319
113.25	37.75	3				.5	.5	1.5	3.25	7.25	8.75	16	26.25	26.75	20.5	18.5	7.75	4.25	.25	.5	142.5	427.5
71	17.75	4							1	5.75	7	4	14.25	13.25	12	11.25	4.5	3.75	.75		77.5	310
22.5	5.5	5					.25	.25	.25	.25	1.5	3	5.5	4.25	5.75	5.25	3.75	2.5	1.5	2	36	180
9	1.5	6					.25	.25	.25	.25	.25	.25	1	2.5	6.5	2.25	2.75	2	1		19.5	117
		7											1.75	.25	4.5	.75	1.25	.75	.25		9.5	66.5
		8											.5		.5	.5	1.5	.75	.25		4	32
		9											1								1	9
(416.5)	Totals	(728.5)	2	4.5	7.5	14.5	45	51.5	92.5	155	178	175	199.5	166	135	82.5	36.5	20	6.5	4.5	(1376)	(1659.5)
			18	36	52.5	87	225	206	277.5	310	178	(1390)	199.5	332	405	330	182.5	120	45.5	36	(1650.5)	

Working Mean x = 67.5 y = 63.5
Class width 1 Inch

The standard deviation of $r^2 = r_1 \times r_2$ has to be estimated by using the general formula for the standard deviation of the product c of two variables a and b;

$$\frac{\sigma_c^2}{c^2} = \frac{\sigma_a^2}{a^2} + \frac{\sigma_b^2}{b^2} + \frac{2R\sigma_a\sigma_b}{ab}$$

R being the correlation coefficient between a and b. Since $-1 < R < +1$, substitution of these limits for R leads to the inequalities

$$\left(\frac{\sigma_a}{a} - \frac{\sigma_b}{b}\right)^2 < \frac{\sigma_c^2}{c} < \left(\frac{\sigma_a}{a} + \frac{\sigma_b}{b}\right)^2 .$$

putting $a = r_1$, $b = r_2$, $c = r^2$ we have

$$\frac{\sigma_{r_1}}{r_1} - \frac{\sigma_{r_2}}{r_2} < \frac{\sigma_{r^2}}{r} < \frac{\sigma_{r_1}}{r_1} + \frac{\sigma_{r_2}}{r_2}$$

Considering the relation $\sigma_r = \dfrac{\sigma_{r^2}}{2r}$

we have $2r(\sigma_{r_1} r_2 - \sigma_{r_2} r_1) < \sigma_r < 2r(\sigma_{r_1} r_2 + \sigma_{r_2} r_1)$
from which we derive with sufficient approximation

$$\sigma_r < \cdot 030$$

A slightly different arrangement for computing r has been made in the following table.

TABLE II

*Correlation between diameter of the stem and length of the lonest flower petal of Trientalis europaea**

	PS	3	15	34	45	30	6	2	0	0	0	0	
PS		−4	−3	−2	−1	0	1	2	3	4	5	6	Total
1	−4	1											1
7	−3	1	4	1	1								7
29	−2	1	9	16	3	1							30
33	−1		2	9	22	9	2	1					45
27	0			8	19	20	4	1					52
8	1	1			7	18	12	6	4				48
1	2				1	8	9	3	2	1			24
	3						3	6	4	1			14
	4							2	2	1	2		7
	5									1	3		4
	6									1		1	2
	Total	4	15	34	53	56	30	19	12	5	5	1	234

* E. Czuber: Die statistischen Forschungsmethoden, Wien, 1921.

TABLE III

x = Diameter of the stem.
y = Length of the longest flower petal in millimeters.
Working mean, $x_m = .825$, $y_m = 34.5$.
Class width of $x = .4$ mm. of $y = 6$ mm.

x	Total times x	P.S. times x	y	Total times y	P.S. times y
−4	16	12	−4	4	4
−3	45	45	−3	21	21
−2	68	68	−2	60	58
−1	53	45	−1	45	33
0	(182)	(170)	0	(130)	(116)
1	30	6	1	48	8
2	38	4	2	48	2
3	36	0	3	42	
4	20	0	4	28	
5	25	0	5	20	
6	6	0	6	12	
	(155)	(10)		(198)	(10)
Mean	−27			+68	

The P.S. columns are the partial sums as explained in the previous table. The work of multiplying the totals by the class marks and of adding them has been separated here from the table.

We obtain $N = 234$, $N_{-x} = 106$, $N_{-y} = 135$

$$r_1 = \frac{170 - 10 - \frac{27}{234} \times 135}{130 + \frac{68}{234} \times 135} = .805$$

$$r_2 = \frac{116 - 10 + \frac{68}{234} \times 106}{182 - \frac{27}{234} \times 106} = .834$$

$$r = .82$$

Pearson's coefficient for this table is $r = .83$.

Finally we illustrate by a small non-grouped table where the partial sums can be written down immediately.

TABLE IV

Correlation between Ages of Husband and Wife

Age of Husband	Age of Wife	Deviation Husband	Deviation Wife
22	18	−8	−8
24	20	−6	−6
26	20	−4	−6
26	24	−4	−2
27	22	−3	−4
27	24	−3	−2
28	27	−2	+1
28	24	−2	−2
29	21	−1	−5
30	25	0	−1
30	29	0	+3
30	32	0	+6
31	27	+1	+1
32	27	+2	+1
33	30	+3	+4
34	27	+4	+1
35	30	+5	+4
35	31	+5	+5
36	30	+6	+4
37	32	+7	+6
Ave 30	26		

Here 0-deviations occur in the third column. Hence[5]

$$\underset{+x}{Sy} = 26 + \tfrac{1}{2} \times 8 = 30, \qquad \underset{+x}{Sx} = 33, \qquad \underset{+y}{Sx} = 31, \qquad \underset{+y}{Sy} = 36,$$

$$r_1 = .86, \quad r_2 = .91, \quad r = .88 \text{ (Pearson's } r = .86)$$

Appendix

Proof of formula (1), page 1. The following notations will be used:

$$(f(x))^0 = \text{probable value of } f(x)$$

$$(f(y))^0_x = \text{probable value of } f(y) \text{ for a fixed } x.$$

$$sgx = \text{sign of } x = \frac{x}{|x|} \text{ for } x \neq 0. \quad sgx = \begin{matrix} +1 \\ 0 \\ -1 \end{matrix} \text{ if } x \gtreqless 0.$$

[5] See page 7.

The assumption of linear regression means that

$$y_x^0 - y^0 = r_{y:x}(x - x^0) \tag{4}$$

We multiply both sides of (4) by some arbitrary function $\phi(x)$ of x and get

$$(y_x^0 - y^0)\phi(x) = r_{y:x}(x - x^0)\phi(x).$$

Both sides are functions of x. We shall take their probable values for all x's.

Now, for a fixed x, $y_x^0\phi(x) = (y\phi(x))_x^0$ and the probable value of $(y\phi(x))_x^0$ for all x's is equal to the total probable value $(y\phi(x))^0$. So we have

$$(y\phi(x))^0 - (y^0\phi(x))^0 = r_{y:x}((x - x^0)\phi(x))^0$$

$$r_{y:x} = \frac{((y - y^0)\phi(x))^0}{((x - x^0)\phi(x))^0} \tag{5}$$

If now we take x^0y^0 as the origin, we get

$$r_{y:x} = \frac{(y\phi(x))^0}{(x\phi(x))^0}$$

and similarly

$$r_{x:y} = \frac{(x\phi_1(y))^0}{(y\phi_1(y))^0}$$

where ϕ_1 is another arbitrary function.

Replacing the probable values by the respective arithmetic means we get

$$r_{y:x} = \frac{Sy\phi(x)}{Sx\phi(x)} \quad \text{and} \quad r_{x:y} = \frac{Sx\phi_1(y)}{Sy\phi_1(y)} \tag{6}$$

with $\bar{x}$, $\bar{y}$ as the origin.

By a suitable choice of the still arbitrary functions ϕ and ϕ_1, we may derive all the various expressions for regression coefficients. Taking, for instance, $\phi(x) = x$, $\phi_1(y) = y$, we get Pearson's expressions. Taking $\phi(x) = sg(x - \alpha_1)$, $\phi_1(y) = sg(y - \alpha_2)$, α_1 and α_2 being constants, we have

$$r_{y:x} = \frac{Sy\, sg(x - \alpha_1)}{Sx\, sg(x - \alpha_1)}, \qquad r_{x:y} = \frac{Sx\, sg(y - \alpha_2)}{Sy\, sg(y - \alpha_2)} \tag{7}$$

and if we make $\alpha_1 = \alpha_2 = 0$

$$r_{y:x} = \frac{Sy\, sg\, x}{Sx\, sg\, x}, \qquad r_{x:y} = \frac{Sx\, sg\, y}{Sy\, sg\, y} \tag{8}$$

Since $Sx = Sy = 0$, we can add Sy or Sx to the numerators and denominators. Adding Sy to the numerator, Sx to the denominator and multiplying both sides of the fraction by $\frac{1}{2}$ we get

$$r_{y:x} = \frac{\frac{1}{2}Sy(sg(x - \alpha_1) + 1)}{\frac{1}{2}Sx(sg(x - \alpha_1) + 1)} \tag{9}$$

Instead of (9) we can write

$$(10)\qquad r_{y:x} = \frac{\mathop{S}\limits_{x>\alpha_1} y + \frac{1}{2}\mathop{S}\limits_{x=\alpha_1} y}{\mathop{S}\limits_{x>\alpha_1} x + \frac{1}{2}\mathop{S}\limits_{x=\alpha_1} x}$$

since the operations of (9) multiply the y ordinates by 0, $\frac{1}{2}$, 1 according as the x's are $\lesseqgtr \alpha_1$.

The expression (10), with a suitable choice of α_1 should be used for the purpose of numerical calculation of r. For instance, when calculating r from the data of Table IV, we took $\alpha_1 = \alpha_2 = 0$ and had

$$r_{y:x} = \frac{\mathop{S}\limits_{+x} y + \frac{1}{2}\mathop{S}\limits_{x=0} y}{\mathop{S}\limits_{+x} x}$$

When dealing with data which are arranged in a grouped table (Tables I and II) we take α_1 equal to the x-ordinate of that classline which is nearest to the mean. $\left(\text{In Table I } \alpha_1 = .5 - \frac{480.5}{1376}\right)$.[6] With that choice of α_1 the sums $\mathop{S}\limits_{x=\alpha_1}$ disappear and the sums $\mathop{S}\limits_{x>\alpha_1}$ are equivalent to the corresponding sums $\mathop{S}\limits_{+x}$. Hence we have

$$(11)\qquad r_{y:x} = \frac{\mathop{S}\limits_{+x} y}{\mathop{S}\limits_{+x} x} \quad \text{and similarly} \quad r_{x:y} = \frac{\mathop{S}\limits_{+y} x}{\mathop{S}\limits_{+y} y}$$

Instead of (9) we can also write

$$(9a)\qquad r_{y:x} = \frac{\frac{1}{2}Sy(sg(x-\alpha_1)-1)}{\frac{1}{2}Sx(sg(x-\alpha_1)-1)}$$

This leads to

$$(11a)\qquad r_{y:x} = \frac{\mathop{S}\limits_{-x} y}{\mathop{S}\limits_{-x} x} \quad \text{and} \quad r_{x:y} = \frac{\mathop{S}\limits_{-y} x}{\mathop{S}\limits_{-y} y}$$

---

[6] It is desirable to chose the absolute values of the α's small so that the maximum number of data enter into the calculation of r. However, to take $\alpha_1 = \alpha_2 = 0$ would necessitate a division of the middle arrays of a grouped table, a laborious process. Hence the choice of the α's as described above.

Proof of the standard deviations of Formula (2).

In my article on standard deviations and correlations of moments[7] the standard deviations of the expressions used in this article have been derived.

In the following, the notation of the Metron article just referred to will be used. We use the symbols:

$$\begin{aligned} P_{m,n} &= \sum x^m \cdot y^n \\ P_{/m,n} &= \sum x^m sgxy^n \\ P_{m,/n} &= \sum x^m y^n sgy \ . \\ P_{/m,/n} &= \sum x^m sgxy^n sgy \end{aligned}$$

The summations indicated extend over all observations. The true or probable values of the same expressions are indicated by using p instead of P.

$$r_{x:y} = r_1 = \frac{P_{1/0}}{P_{0/1}}$$

We derive the standard deviations by defining the deviations as first variations.

$$\log r_1 = \log P_{1/0} - \log P_{0/1}$$

$$\frac{\delta r_1}{r_1'} = \frac{\delta P_{1/0}}{p_{1/0}} - \frac{\delta P_{0/1}}{p_{0/1}}$$

$$\sigma r_1^2 = [(\delta r_1)^2]^0 = (r_1')^2 \left[\left(\frac{\delta P_{1/0}}{p_{1/0}} - \frac{\delta P_{0/1}}{p_{0/1}}\right)^2\right]^0 \tag{12}$$

The probable values of the terms on the right hand side of the last equation are derived on pages 17–19 and listed on pages 32–33 of the Metron article referred to. The proofs which imply essentially a process of variation of Stieltje's integrals will not be given here. From pages 32–33 we take

$$\begin{aligned} &[(\delta P_{1/0})^2]^0 = \frac{p_{20} - p_{1/0}^2}{N}, \qquad [(\delta P_{0/1})^2]^0 = \frac{p_{02} - p_{0/1}^2}{N} \\ &[(P_{1/0}\delta P_{0/1}]^0 = \frac{p_{11} - p_{1/0}p_{0/1}}{N} \end{aligned} \tag{13}$$

so that

$$\sigma_{r_1}^2 = \frac{1}{N}(r_1')^2 \left[\frac{p_{20}}{p_{1/0}^2} + \frac{p_{02}}{p_{0/1}^2} - \frac{2p_{11}}{p_{1/0}p_{0/1}}\right] \tag{14}$$

Assuming Gaussian distribution, we can put

$$p_{20} = \frac{\pi}{2} p_{/10}^2 \qquad p_{02} = \frac{\pi}{2} p_{0/1}^2 \qquad p_{11} = r\sqrt{p_{02}p_{20}} = r\frac{\pi}{2}p_{/10}p_{0/1}$$

---

[7] Felix Bernstein: "Die mittleren Fehlerquadrate und Korrelationen der Potenzmomente und ihre Anwendung auf Funktionen der Potenzmomente," Metron, Vol. X, N. 3 Nov. 1932).

Hence

$$\sigma^2_{r_1} = \frac{1}{N}\cdot\frac{\pi}{2}(r_1')^2\left(1 + \frac{p^2_{/10}}{p^2_{1/0}} - 2r\frac{p_{/10}}{p_{1/0}}\right) \tag{15}$$

Replacing the theoretical values by their corresponding empirical values, we have

$$\sigma^2_{r_1} = \frac{\pi r_1^2}{2N}(1 + m^2 - 2rm) \qquad \text{where } m = \frac{Sx\,sg\,x}{Sx\,sg\,y} \tag{16}$$

The formula for $\sigma^2_{r_1}$ has been derived here for the value of r_1 as given by (8) i.e. $r_1 = \dfrac{Sx\,sg\,y}{Sy\,sg\,y}$. In fact, we used $r_1 = \dfrac{Sx\,sg\,(y-\alpha)}{Sy\,sg\,(y-\alpha)}$ in the examples in the article, and α had some value absolutely smaller than .5. To use equation (16) for the standard deviation of r_1 is within the limits of the required degree of accuracy; hence we shall disregard the difference. In a later paper the standard deviation of r_1 for any α will be derived by using the method described in the Metron article, for a different purpose.

To prove the statement in the footnote to page 7

To find the value of r_2 that makes

$$Sf(x)\,(y - r_2x)^2 \text{ a minimum.}$$

By differentiating we get

$$Sf(x)(y - r_2x)\,x = 0$$

$$r_2 = \frac{Sxf(x)y}{Sxf(x)x}$$

If $f(x) = 1$ we get Pearson's coefficient.

If $f(x) = \dfrac{1}{|x|}$ $(x \neq 0)$ we get

$$r_2 = \frac{S\dfrac{x}{|x|}y}{S\dfrac{x}{|x|}x} = \frac{Sy\,sg\,x}{Sx\,sg\,x}$$

New York University,
Departments of Anatomy of the Graduate School and the College of Dentistry.

METHODS OF OBTAINING PROBABILITY DISTRIBUTIONS[1]

By Burton H. Camp

The emphasis of this paper will be on method. Special results will be cited in order to illustrate the methods rather than to summarize achievement in the field; for that has been done already by Rider (1930, 1935) Irwin (1935) and Shewhart (1933) in recent surveys. The purpose is to describe and to illustrate most of the methods that have been used to determine exact probability distributions, and to show that they are all derivable from one fundamental theorem. In order to prove this unity in a simple manner, it will be desirable to omit from consideration methods which are essentially ingenious forms of counting, such as are used in sampling without replacements from finite universes, and in finding the sampling distribution of a percentile.

The general problem to be discussed may be stated as follows: N individuals $(t_1, \cdots, t_N)$ are drawn, one at a time with replacements, from a universe whose probability distribution is $\phi(t)$. A certain single valued function of the t's is formed. This is called a parameter of the sample, and is frequently also, but not necessarily, a useful estimate of the corresponding parameter of the universe. The problem is to find its probability distribution, $f(x)$. As usual, a probability distribution is a function which is required to be defined, except perhaps at a set of measure zero, throughout the infinite domain of its variables; it is nowhere negative, and its integral over its domain is unity.

Most of the more recent developments of the theory relate to a more general form of this problem. Instead of N individuals, there are N sets of n individuals in each set, and these sets are drawn respectively from $M(M \leqq N)$ universes, each of which is described by a function of n independent variables, thus:

$$\phi^{(i)}(t_1, \cdots, t_n); \; (i = 1, \cdots, M). \tag{1}$$

Instead of a single parameter there are P parameters, and each is a single valued function of the observed values of the nN individuals in the sample, thus:

$$x_i = g_i(t_1^{(1)}, \cdots, t_n^{(1)}; \cdots; t_1^{(N)}, \cdots, t_n^{(N)}); \; (i = 1, \cdots, P) \tag{2}$$

The first method to be described is fundamental and will be designated as

Theorem I. Let it be required that each g as described in (2) be not only single valued but also constant at most in a set of measure zero in the nN-way space of the t's. Then

$$\int_p f(x_1, \cdots, x_P)\, dX = \int_q \phi(t_1^{(1)}, \cdots, t_n^{(N)})\, dT \tag{I}$$

[1] Presented to the American Mathematical Society at a meeting devoted to expository papers on the theory of statistics, April 11, 1936.

where X is the space of x's and T the space of the t's, p is any measurable set of points in X, and q is the set in T for which g is in p. Often p is the P dimensional cube $(x_i + \Delta x, i = 1, \cdots P)$ at the point $(x_1, \cdots, x_p)$ and then q is the set where

$$x_i \leqq g_i \leqq x_i + \Delta x; \qquad (i = 1, \cdots, P) \tag{3}$$

and ϕ is the simultaneous distribution of the sets of t's,

$$\phi^{(1)}(t_1^{(1)}, \cdots, t_n^{(1)}) \cdots \phi^{(N)}(t_1^{(N)}, \cdots, t_n^{(N)}). \tag{4}$$

In this $\phi^{(i)}$ is the universe from which the $t^{(i)}$ set of t's is drawn. Obviously, if $N > M$, some of the $\phi^{(i)}$'s are identical, and then it is assumed that the several sets are drawn independently. Often, all of the N sets of t's are drawn from the same universe. Then $M = 1$ and all these ϕ's are identical, and (4) becomes

$$\phi = [\phi^{(1)}(t_1^{(1)}, \cdots, t_n^{(1)}] \cdots [\phi^{(1)}(t_1^{(N)}, \cdots, t_n^{(N)})].$$

In the special case where there is but one parameter ($P = 1$) and but one individual in the sample ($n = N = 1$), and p is an interval, formula (I) becomes

$$\int_x^{x+\Delta x} f(x)\, dx = \int \phi dt; \tag{Ia}$$

and in the very special case where it is also true that q is an interval it becomes

$$f(x) = \phi(t) \cdot \left| \frac{dt}{dx} \right|, \tag{Ib}$$

provided also that certain derivatives (to be specified later in the proof) exist, where t is now the inverse solution of the equation,

$$x = g(t). \tag{5}$$

The proof of formula (I) is immediate, if one is willing to assume the existence of the probability distribution f; for then the left side is by definition the probability that the x's lie in p, and this is also the meaning of the right side of (I). (Ia) can be proved without assuming initially the existence of $f(x)$, for then the existence of $f(x)$ can be inferred from the existence of the right side of (Ia), because $f(x)$ may be set equal (except perhaps at a set of measure zero) to the upper right hand derivative, with respect to Δx (Δx is a variable, and x is fixed), of $\int_q \phi\, dt$, provided that one adds the condition that this derivative is nowhere infinite. The point at issue here is merely the existence of a primative for a monotone increasing function of Δx. (Ib) may be derived from (Ia) by taking the derivative of both sides with respect to Δx, if the derivatives are continuous.

Theorem I, in these various forms is used a great deal, especially in the last form (Ib). This affords one freedom to choose the most desirable function for purposes of tabulation. R. A. Fischer's z distribution, a logarithm, is an important illustration. Many authors have been interested in so choosing the

function that its distribution shall be normal. They include several of the older writers, and more recently H. L. Rietz (1921, 1927), and G. A. Baker (1932, 1934). However, the theorem is of special importance in the theory, for all the other principal methods of obtaining probability distributions are essentially corollaries of it. These corollaries will be called Theorems II, III, and IV.

THEOREM II. Let $\bar{p}$ (the measure of p) and $\bar{q}$ (the measure of q) be infinitesimals of the same order and let both the oscillation of f(*i.e.* maximum f- minimum f) in p and the oscillation of ϕ in q be infinitesimals; then (I) may be written,

$$\text{(II)} \qquad f\bar{p} = \phi\bar{q},$$

where f applies to any point of p and ϕ to the corresponding point of q. This equation (II) is an approximate equation in the sense that differences of higher order than those retained are neglected. In particular, with the conditions used in formula (Ia), equation II becomes

$$f\Delta x = \phi\bar{q}.$$

The left side of (II) is an approximation to the probability sought. The right side shows that, in order to evaluate it, one need only find the volume in T space of the differential element q and multiply it by the value of ϕ in q. Formula (II) expresses the so-called geometrical method used by many authors, *e.g.*, by R. A. Fisher (1915, 1925), by Wishart (1928), and by Hotelling (1925, 1927). The chief difficulty in connection with it is in finding the volume of nN-dimensional q. In order to display the advantages and disadvantages of this method we shall pause at this point and look at a concrete example.[2]

Let two individual (t_1, t_2) be drawn independently from a normal universe and consider the simultaneous distribution $f(x, y)$ of the sum, $x = t_1 + t_2$, and product, $y = t_1t_2$, the mean of the universe being chosen as the origin. Here $N = 2$, $n = 1$, $M = 1$, and so,

$$\text{(6)} \qquad \phi = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}(t_1^2 + t_2^2)} = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}(x^2 - 2xy)}.$$

The point set q is the area lying between the two adjacent hyperbolae,

$$t_1t_2 = y, \qquad t_1t_2 = y + \Delta y,$$

and also between the two adjacent lines,

$$t_1 + t_2 = x, \qquad t_1 + t_2 = x + \Delta x,$$

where Δx and Δy are infinitesimals and are equal. This area may be computed by simple integration and is:

---

[2] See also C. C. Craig (1936). Craig uses another method to be explained later (formula IIIa).

$$q = \frac{2\Delta x\, \Delta y}{\sqrt{x^2 - 4y}} \qquad \text{if} \quad x^2 > 4y,$$

$$= 0 \qquad \text{if} \quad x^2 < 4y.$$

Hence II gives us immediately the desired result:

$$f(x, y)\, \Delta x \Delta y = \frac{1}{\pi\sigma^2}\, e^{-\frac{x^2-2y}{2\sigma^2}} \frac{1}{\sqrt{x^2 - 4y}} \cdot \Delta x \Delta y, \qquad \text{if} \quad x^2 > 4y,$$

$$= 0 \quad \text{if} \quad x^2 < 4y.$$

If $x^2 = 4y$, $\bar{q}$ is an infinitesimal of lower order than $\bar{p} = (\Delta x)^2$, and so Theorem II does not apply. In this case we must go back to Theorem I, and from that we can learn that the probability,

$$\int_p f\, dx\, dy,$$

is an infinitesimal of the first order if $p = \Delta x\, \Delta y = (\Delta x)^2$ is of the second order. Hence it cannot be approximately represented by a finite number times $\bar{p}$. The oscillation of f in p is infinite. The form of the surface $f(x, y)$ is interesting. The ordinates rise to infinity on the contour of the parabola $x^2 = 4y$, and vanish within it. The surface is symmetrical with respect to the plane $x = 0$, but not with respect to the plane $y = 0$. However, it is clear that the total probability of any given product, y (*i.e.* the probability of this y for all possible values of x), is the same as the total probability of $-y$; hence

$$\int_{-\infty}^{\infty} f(x, y)\, dx = \int_{-\infty}^{\infty} f(x, -y)\, dx,$$

and the corresponding formulae,

$$\frac{2}{\pi\sigma^2}\, e^{\frac{y}{\sigma^2}} \int_{\sqrt{4y}}^{\infty} e^{-\frac{x^2}{2\sigma^2}} \frac{1}{\sqrt{x^2 - 4y}}\, dx \qquad (y > 0),$$

and

$$\frac{2}{\pi\sigma^2}\, e^{\frac{y}{\sigma^2}} \int_{0}^{\infty} e^{-\frac{x^2}{2\sigma^2}} \frac{1}{\sqrt{x^2 - 4y}}\, dx \qquad (y < 0),$$

must be equal; both may be reduced to the single form

$$F(y) = \frac{1}{\pi\sigma^2} \int_0^{\infty} e^{-\frac{1}{2\sigma^2}\left(t^2 + \frac{y^2}{t^2}\right)} \frac{dt}{t}, \qquad \text{if} \quad y \neq 0.$$

This is the probability distribution of y.

With this example before us, let us now reconsider the theory:

(*i*) The requirement (in II) that the oscillation of ϕ be infinitesimal in q

will be satisfied if one can show that ϕ may be expressed as a continuous function of the parameters $(x_1, x_2, \cdots, x_P)$. In our example these parameters were x and y and ϕ was so expressible (6). But if we had tried initially to find by means of (II) the distribution of the product y, independently of what values x might have, we should have been stopped at this point, because ϕ is not expressible in terms of y alone. We should also have been stopped by the requirement that $\bar{q}$ be infinitesimal of order Δy, for q would have been the space between two hyperbolas and its area for any fixed $(\Delta y > 0)$ would have been infinite. But, when thus stopped at that first point, it would have been clearly indicated to us that the distribution of y might have been found *via* the detour of finding the simultaneous distribution of both x and y, because an attempt to express ϕ in terms of y would have led to the given expression in terms of both x and y. For a similar reason R. A. Fisher (1925) was able to find the distribution of the variance by finding first the simultaneous distribution of the variance and the mean. Also, he was thus able to find the distribution of the coefficient of correlation by finding first the simultaneous distribution of all the first and second order moments.

(*ii*) A distinct advantage of this method is that q is independent of the universe ϕ, so that once found it may be used in connection with any universe which satisfies the condition that it can be expressed as a continuous function of the parameters. Thus, the distribution of the sum and product in our example may equally well be found for the universe described by the Type III curve, $Ate^{-at}(t > 0)$. For, then

$$\phi = A^2\, t_1\, t_2\, e^{-a(t_1+t_2)} = A^2\, y\, e^{-ax},$$

and so, using one-half of the same $\bar{q}$ as before, since now $x, y \geqq 0$,

$$f(x, y) = A^2\, y\, e^{-ax} \frac{2}{\sqrt{x^2 - 4y}} \qquad \text{if} \quad x^2 > 4y,$$

$$= 0 \qquad \text{if} \quad x^2 < 4y.$$

From this, $F(y)$ can be found by integration (*c.f.* Kullbach, 1934)

$$F(y) = A^2 y \int_{\sqrt{4y}}^{\infty} \frac{e^{-ax}}{\sqrt{x^2 - 4y}}\, dx = \frac{A^2 y}{2} \int_0^{\infty} \frac{e^{-a\left(u+\frac{y}{u}\right)}}{u}\, du.$$

As another illustration, consider a normal universe of n intercorrelated variables in which all the total intercorrelations are equal to r (*e.g.*, the statures of n brothers) and let the sample be a single group of n (one individual for each variable).

$$\phi = \frac{1}{(2\pi)^{n/2} R}\, e^{-\frac{1}{2R}\left[k_1 \Sigma t_i^2 + k_2 \sum_{i \neq j} t_i t_j\right]},$$

where $R = (1 - r)^{n-1}[1 - (n - 1)r]$, $k_1 = (1 - r)^{n-2}[1 - (n - 2)r]$, and $k_2 = -r(1 - r)^{n-2}$. Suppose one wishes to find the simultaneous distribution

of the variance x and the mean y for such samples.[3] Since for Student's problem Fisher has found the value of q for this x and y to be

$$\bar{q} = cx^{\frac{n-3}{2}} \Delta x \Delta y,$$

their distribution $f(x, y)$ for this universe may be written down immediately. In terms of x and y the bracket in the exponent of ϕ is $y^2(k_1 n - k_2 n + k_2 n^2) + xn(k_1 - k_2)$, and so $f(x, y)$ is the product of $\bar{q}$ and this form of ϕ:

$$f(x, y) = K e^E x^{\frac{n-3}{2}}, \qquad E = -\frac{1}{2R}[(k_1 n - k_2 n + k_2 n^2)y^2 - n(k_1 - k_2)x].$$

(*iii*) Another attribute of this method is that it sometimes lends itself to easy extensions from a simple case where there is only one restriction ($N - 1$ degrees of freedom) to similar cases when there are more restrictions. Thus R. A. Fisher (1924) proceeded from the variance of a sample from a single universe to the variance from a set of universes, as required in the theory of analysis of variance; and thus also (1915) he had proceeded from the distribution of r to that of multiple R; and Hotelling (1927) showed how these distributions could be obtained when the values of each variate were themselves intercorrelated (as in a time series) and not merely correlated with values of the other variates.

THEOREM III. Now let us consider again the fundamental form (I). For convenience let $nN = m$. If the conditions will not permit us to write the right side in the form in (II), it is still possible that we may be able to find that $(m + 1)$-dimensional volume by some other method. In particular, whenever it is possible to iterate the integral once we have the formula:

$$\text{(III)} \qquad \int_p f dX = \int_{T'} dT' \int_{q_m} \phi \, dt_m,$$

where q_m is the section of q by t_m space at the point $(t_1, \cdots, t_{m-1})$ of T' space, T' space being the space of the $(t_1, \cdots, t_{m-1})$ coordinates. With added conditions one may deduce from (III), for the case where there is but a single parameter x, the approximate equation:

$$\text{(IIIa)} \qquad f\, dx = dx \int_{T'} dT' \cdot \phi(t_1, \cdots, t_m) \frac{dt_m}{dx},$$

in which t_m is supposed to have been expressed in terms of the other coordinates by solving the equation $x = g(t_1, \cdots, t_m)$. It is an approximate equation in the same sense as (II) was. Sufficient conditions for this change in the left side of (III) have already been mentioned in discussing (II). The propriety of making the corresponding change in the right hand side may be left for determination when the form of ϕ is given. It will perhaps be sufficient her to point out that our earlier example illustrates both the case where this change

[3] A special case of a more general problem solved first by R. A. Fisher.

is permissible and where it is not. For, let it be required to find the distribution $f(y)$ of the product $y = t_1 t_2$ without reference to the sum, $t_1 + t_2$. Formula (III) yields

$$(7) \qquad \int_y^{y+\Delta y} f(y)\, dy = 2 \int_0^\infty dt_1 \int_{y/t_1}^{(y+\Delta y)/t_1} dt_2 \cdot \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}\left(t_1^2 + t_2^2\right)}.$$

This is valid for every value of y including $y = 0$. If $y \neq 0$, we may change the right hand side as in (IIIa) and obtain as the probability that y is in the interval $(y, y + \Delta y)$:

$$(8) \qquad \int_y^{y+\Delta y} f(y)\, dy = \frac{\Delta y}{\pi\sigma^2} \int_0^\infty \frac{1}{t_1} e^{-\frac{1}{2\sigma^2}\left(t_1^2 + \frac{y^2}{t_1^2}\right)} dt_1 + \epsilon,$$

where ϵ is a differential of higher order than Δy. This may be proved by computing the difference between the value of (7) when t_2 has constantly the value $(y + \Delta y)/t_1$ and when it has constantly the value y/t_1. If $y = 0$ this change in the right side of (7) is not valid; it is easily seen that in this case the integral on the right of (8) is infinite. It may be shown, however, in this case that

$$(9) \qquad \int_0^{\Delta y} f(y)\, dy = \tfrac{1}{4} - \frac{1}{2\pi} \int_1^\infty \frac{e^{-\frac{x\Delta y}{\sigma^2}}\, dx}{x\sqrt{x^2 - 1}},$$

and that this is an infinitesimal, and that it is of order as small as one.

Many authors think of (IIIa) as the fundamental formula in the theory of probability distributions. One of the simplest and earliest applications of it was to establish the so-called reproductive property of the normal law: that the sum of two variates is distributed normally if each is distributed normally. Jackson (1935) has used it to establish a similar property for two Type III distributions which have the same exponent of e. Usually this integral is difficult to evaluate when $N > 2$ because of the unsymmetrical form into which it is cast, but when $N = 2$ and there is but one parameter (IIIa) it is perhaps the most convenient of all the formulae.

Theorem IV. An exceedingly useful formula is obtainable from (I) in the following manner. Let $\theta(x_1, \cdots, x_P; \alpha_1, \cdots, \alpha_Q)$ be a finite single valued function of the old parameters (x) and of some new parameters (α). Subject to general conditions to be stated we may write:

$$(\mathrm{IV}) \qquad \int_X \theta f\, dX = \int_T \theta' \phi\, dT,$$

an identity with respect to each α, where θ' is the result of substituting (2) for the x's in θ.

Since this theorem has not been proved in this general form, an outline of the proof will be given. Sufficient conditions are:

(a) All the integrals involved shall exist.

(*b*) If p is limited (in the sense that it lies within a finite hypersphere), so is q, and conversely.

Proof. Let X_0 be a limited p set and T_0 the corresponding q set such that both (*c*) and (*d*) hold ($\epsilon > 0$):

$$(c) \qquad \left| \int_{X_0} f\theta \, dX - \int_X f\theta \, dX \right| < \epsilon,$$

$$(d) \qquad \left| \int_{T_0} \phi\theta' \, dT - \int_T \phi\theta' \, dT \right| < \epsilon.$$

It is easy to see that such an X_0 and a corresponding T_0 do exist, as follows: Let X_0' be a limited set for which (*c*) is true, and for which it will remain true no matter what points are added to X_0'. Similarly, let T_0' be a limited set for which (*d*) is true and for which it will remain true, no matter what points are added to T_0'. Presumably X_0' and T_0' do not correspond to each other, but we may now let X_0 be the totality of all the points of X_0' and of all those points of X corresponding to T_0', and let T_0 be the totality of all the points of T_0' and of all those points of T corresponding to X_0'. Then X_0 and T_0 do correspond to each other and have the desired properties (*c*) and (*d*). Now, since θ is finite, it is limited in X_0. Let

$$(e) \qquad | \theta | < H \text{ in } X_0.$$

Divide the interval $(-H, H)$ into s equal subintervals of length h, thus defining in X_0 according to Lebesgue the measurable sets,

p_i $(i = 1, \cdots, s)$, and corresponding q_i sets in T_0:

$$(f) \qquad \begin{cases} Os\ \theta \leqq h \text{ in } p_i, \\ Os\ \theta' \leqq h \text{ in } q_i. \end{cases}$$

Choose arbitrarily any point of p_i and let k_i be the corresponding value of θ. Then let

$$\bar{\theta} = k_i \text{ in } p_i \ (i = 1, \cdots, s), \text{ and similarly let}$$

$$\bar{\theta}' = k_i \text{ in } q_i \ (i = 1, \cdots, s).$$

Then

$$\int_{X_0} \bar{\theta} f \, dX = \sum_i k_i \int_{p_i} f \, dX,$$

and

$$\int_{T_0} \bar{\theta}'\phi \, dT = \sum_i k_i \int_{q_i} \phi \, dt.$$

Since by (I)

$$\int_{p_i} f \, dX = \int_{q_i} \phi \, dT,$$

$$(g) \qquad \int_{X_0} \bar{\theta} f \, dX = \int_{T_0} \bar{\theta}'\phi \, dT.$$

Now

$$\left| \int_{X_0} (\bar{\theta} - \theta) f \, dX \right| \leqq \int_{X_0} | \bar{\theta} - \theta | f \, dX \leqq h \int_{X_0} f \, dX,$$

and

$$\left| \int_{T_0} (\bar{\theta}' - \theta') \, dX \right| \leqq h \int_{T_0} \phi \, dX.$$

So, as h approaches zero both sides of (g) approach limits and their limits are equal:

$$\int_{X_0} \theta f \, dX = \int_{T_0} \theta' \phi \, dT.$$

Hence by (c) and (d) the integrals

$$\int_X \theta f \, dx, \qquad \int_T \theta' \phi \, dT,$$

differ at most by 2ϵ, and so, being independent of ϵ they do not differ at all.

In order to determine the form of f from (IV) one must first evaluate the right side,

$$\int_T \theta\phi \, dt = \psi(\alpha_1, \cdots, \alpha_q);$$

and then solve the integral equation,

$$\int_X \theta f \, dX = \psi. \tag{10}$$

It is the solution of this equation that usually presents the most difficulty. Particular forms of θ that are being used are

$$\theta = e^{\alpha_1 x + \cdots + \alpha_P x_P}, \tag{11}$$

in which case ψ is said to be the "characteristic function" or "moment generating function"; and

$$\theta = x_1^{\alpha_1} \cdots x_P^{\alpha_P}, \tag{12}$$

in which case ψ is a "moment function" or "moment" of f. Other forms might be used. For example, a very convenient method of demonstrating the correctness of the usual formula for the simultaneous distribution of the correlation (x), means (y, z), and variances (u, v), in samples from a normal bivariate universe is by the use of

$$\theta = e^{\alpha_1(u^2 + v^2 + y^2 + z^2) + \alpha_2(uvx + yz)}.$$

This method of finding f is not a final determination of the probability function desired until it has been shown that the solution is unique, a serious problem

in itself; it is one of those which Professor Shohat may consider.[4] There are three methods of solving the integral equation (10):

(*i*) The first might be called guessing. Though unscientific, it is in fact often effective. Especially is it available if the distribution has already been surmised but not demonstrated. Thus, it was open to Student (1908) when he correctly surmised the distribution of the variance. Similarly it was open to Soper (1913) when he incorrectly surmised the distribution of r.

(*ii*) Papers by Romanovsky (1925) and Wilks (1932) have shown how the problem of solving the integral equation may be shifted to the problem of solving a partial differential equation, but this in turn may involve the solution of another equally difficult integral equation in the process or determining the arbitrary function.

(*iii*) If each α be replaced by an imaginary βi and one uses a Fourier transform, one arrives at a set of formulae which are most important. For the case where there is but one x and one β, they may be written:

$$\int_{-\infty}^{\infty} e^{i\beta x} f(x)\,dx = \int_T e^{i\beta g}\,\phi\,dT = \psi(\beta). \tag{13}$$

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\beta x}\,\psi(\beta)\,d\beta. \tag{14}$$

Dodd (1925) has given an equivalent set of formulae involving only real variables. It is easy to prove that both sets may be changed to the single formula,

$$f(x) = \frac{1}{\pi}\int_T \phi\,dt \int_0^{\infty} \cos\beta(x - g)\,d\beta. \tag{15}$$

Kullbach (1936) has established the validity of the formulae corresponding to (13) and (14) for the general case of $(P + Q)$ parameters. Wishart and Bartlett (1933) used the general forms to find the distribution of the generalized product moment in samples from an n-dimensional normal system.

When the solution of the integral equations of (IV) cannot be found, one has to put up with the semi-invariants or with the moments of f. Formulae (IV) and (11) yield the semi-invariants, (IV) and (12) the moments about the given origin, and from either of these one may obtain the moments about the mean point. These methods are old but they are still important. Time does not permit me to discuss them, because it would not be proper to close this paper without some reference to limit methods.

Limit Methods. It is well known that the distribution of means of samples taken from almost[5] any universe approaches the normal law as a limit as N becomes infinite. This theorem is subject to great generalizations, as is indicated in papers of A. Liapounoff (1901), S. Bernstein (1926), Romanovsky

[4] In a later paper at the same symposium.

[5] There are exceptions. *E. g.*, means of samples taken from the universe $a/\pi(a + t^2)$ have a distribution identical with the universe itself.

(1929, 1930) and C. C. Craig (1932). Subject to very general conditions it has been shown that: If the characteristic function of one probability distribution contains a parameter and approaches as a limit, uniformly in every finite domain of its variables, the characteristic function of another probability distribution; then the first distribution approaches as a limit the second distribution. Hence S. Bernstein and Romanovsky have shown that: If the universe is an n-way correlation solid of a certain very general type, then the n means obtained by a selection of a sample of N sets of variates, $x_i = \frac{1}{N}(t_{i_1} + \cdots + t_{iN})$, $(i = 1, \cdots, n)$, have a distribution which approaches as a limit a normal correlation solid as N becomes infinite. A similar theorem has been established also in the interesting case of Romanovsky's "belonging coefficients", which include K. Pearson's coefficient of racial likeness. Also, by the method of maximum likelihood, Hotelling (1930) has proved that under certain general conditions all optimum estimates of the parameters of a frequency distribution have a joint distribution approaching the normal as N becomes infinite. The validity of the method of maximum likelihood when used for this purpose has been established by J. L. Doob (1934).

Finally, one may note an apparently new limit theorem of another type. Its general nature will be obvious from the following application:

Let a sample of N be drawn from the universe,

$$\phi = Ae^{-at^{2\lambda}}, \quad \text{if} \quad t > 0,$$
$$= 0 \quad \text{if} \quad t \leqq 0.$$

It is readily proved, by means of (IV), that the distribution $f(x)$ of the parameter,

$$x = (t_1^{2\lambda} + \cdots + t_N^{2\lambda})^{1/N}$$

is a curve of the form,

$$f(x) = Bx^{N-1} e^{-x^{2\lambda}} \text{ where } x > 0,$$
$$= 0 \quad \text{elsewhere.}$$

Now let λ become infinite. The universe approaches as a limit the rectangle:

$$\Phi = A \text{ where } 0 \leqq t < 1,$$
$$= 0 \quad \text{elsewhere.}$$

The parameter x approaches as a limit X, where X = maximum t_i. The distribution $f(x)$ approaches as a limit the new distribution,

$$F(X) = NX^{N-1} \text{ where } 0 < |X| < 1,$$
$$= 0 \quad \text{elsewhere.}$$

Hence we have proved in a new way, what was already known: that the distribution of the greatest variate obtained by sampling from a rectangular universe is of the form $F(X)$.

The limit theorem implicit in this illustration can be established in sufficient generality, but I do not yet know whether it has other applications of value.

REFERENCES

G. A. Baker, Transformations of bimodal distributions, Annals of Mathematical Statistics, vol. 1 (1932), pp. 334-344.

G. A. Baker, Transformation of non-normal frequency distributions into normal distributions, Annals of Mathematical Statistics, vol. 5 (1934), pp. 113-123.

S. Bernstein, Su. l'extension du théorème limite du calcul des probabilités aux sommes de quantités dependantes, Mathematische Annalen, vol. 97 (1926), pp. 1-59.

C. C. Craig, On the composition of dependent elementary errors, Annals of Mathematics, vol. 33 (1932), pp. 184-206.

C. C. Craig, On the frequency function of xy, Annals of Mathematical Statistics, vol. 7 (1936), pp. 1-15.

E. L. Dodd, The frequency law of a function of variables with given frequency laws, Annals of Mathematics, vol. 27 (1925), pp. 12-20.

J. L. Doob, Probability and statistics, Transactions of the American Mathematical Society, vol. 36 (1934), pp. 759-775.

R. A. Fisher, Frequency distributions of the values of the correlation coefficient in samples from an indefinitely large population, Biometrika, vol. 10 (1915), pp. 507-521.

R. A. Fisher, On a distribution yielding the error function of several well known statistics, Proceedings of the International Mathematical Congress, Toronto, 1924, vol. 2, pp. 805-813.

R. A. Fisher, Applications of "Student's" distribution, Metron, vol. 5, no. 3, (1925), pp. 90-104.

H. Hotelling, The distribution of correlation ratios calculated from random data, Proceedings of the National Academy of Sciences, vol. 11, no. 10 (1925), pp. 657-662.

H. Hotelling, An application of analysis situs to statistics. Bulletin of the American Mathematical Society, vol. 33 (1927), pp. 467-476.

H. Hotelling, The consistency and ultimate distribution of optimum statistics, Transactions of the American Mathematical Society, vol. 32 (1930), pp. 847-859.

J. O. Irwin, Recent advances in mathematical statistics, Journal of the Royal Statistical Society, vol. 98, part 1 (1935), pp. 88-92.

D. Jackson, Mathematical principles in the theory of small samples, American Mathematical Monthly, vol. 42 (1935), pp. 344-364.

S. Kullbach, An application of characteristic functions to the distribution problem of statistics, Annals of Mathematical Statistics, vol. 5 (1934), pp. 263-307.

S. Kullbach, On certain distribution theorems of statistics, Bulletin of the American Mathematical Society, vol. 42 (1936), pp. 407-410.

A. Liapounoff, Nouvelle forme du théorème sur la limite de probabilité, Memoires de l'Académie de St. Pétersbourg (8), vol. 11 (1901).

P. R. Rider, A survey of the theory of small samples, Annals of Mathematics, vol. 31 (1930), pp. 577-628.

P. R. Rider, Recent progress in statistical method, Journal of the American Statistical Association, vol. 30 (1935), pp. 58-88.

H. L. Rietz, Frequency distributions obtained by certain transformations of normally distributed variates, Annals of Mathematics (2) vol. 23 (1921-22), pp. 292-300.

H. L. Rietz, On certain properties of frequency distributions of the powers and roots of the variates of a given distribution, Proceedings of the National Academy of Sciences, vol. 13 (1927), pp. 817–820.

U. Romanovsky, On the moments of the standard deviations and of correlation coefficients in samples from a normal population, Metron, Vol. 5(4) (1925), pp. 1–45.

U. Romanovsky, Sur une extension du théorème de A. Liapounoff sur la limite de probabilité, Bulletin de l'Academie des Sciences de l'U. S. S. R. (1929), pp. 209–225.

U. Romanovsky, On the moments of means of functions of one or more random variables, Metron, vol. 8(1–2) (1930), pp. 251–291.

W. A. Shewhart, Annual Survey of Statistical Technique: Sample Theory, Econometrica, vol. 1 (1933), pp. 225–237.

H. E. Soper, On the probable error of the correlation coefficient to a second approximation, Biometrika, Vol. 9 (1913), pp. 91–115.

Student, The probable error of the mean, Biometrika, vol. 6 (1908), pp. 1–25.

S. S. Wilks, On the distributions of statistics in samples from a normal population of two variables with matched sampling of one variable, Metron, vol. 9(3–4) (1932), pp. 87–126.

J. Wishart, The generalized product moment distribution in samples from a normal multivariate population, Biometrika, vol. 20A, (1928), pp. 32–52.

J. Wishart and M. S. Bartlett, The generalized product moment distribution in a normal system, Proceedings of the Cambridge Philosophical Society, vol. 29 (1933), pp. 260–270.

Wesleyan University.

MOMENT RECURRENCE RELATIONS FOR BINOMIAL, POISSON AND HYPERGEOMETRIC FREQUENCY DISTRIBUTIONS[1]

By John Riordan

1. **Introduction.** This paper gives the development of recurrence relations for moments about the origin and mean of binomial, Poisson, and hypergeometric frequency distributions from the basis of the moment arrays defined by H. E. Soper.[2] This procedure has the advantage of expressing the moments in terms of coefficients which are alike for the three distributions and are derivable by a single process, thus providing a degree of formal coordination of the distributions. For both kinds of moments, the coefficients satisfy relatively simple recurrence relations, the use of which leads to recurrence relations for the moments, thus unifying the derivation of these relations for the three distributions. The relations derived in this way for the hypergeometric distribution are apparently new. Apparently new recurrence relations for certain auxiliary coefficients in the expression of the moments about the mean of binomial and Poisson distributions are also given.

This course of development involves repetition of a number of well-known results which is justified, it is hoped, by the unification obtained.[3]

---

[1] Presented to the American Mathematical Society, Sept. 3, 1936.

[2] *Frequency Arrays*, Cambridge, 1922.

[3] The following bibliography is taken from a paper *On the Bernoulli Distribution*, Solomon Kullback, Bull. Am. Math. Soc., **41**, 12, pp. 857–864, (Dec., 1935):

A. Fisher, *The Mathematical Theory of Probabilities*, 2d ed., p. 104 ff.

H. L. Rietz, *Mathematical Statistics*, 1927, p. 26 ff.

V. Mises, *Wahrscheinlichkeitsrechnung*, 1931, pp. 131–133.

Risser and Traynard, *Les Principes de la Statistique Mathématique*, 1933, pp. 39–40 and 320–321.

V. Romanovsky, *Note on the moments of the binomial* $(q + p)^n$ *about its mean*, Biometrika, vol. 15 (1923), pp. 410–412.

A. T. Craig, *Note on the moments of a Bernoulli distribution*, Bull. Am. Math. Soc., vol. 40 (1934), pp. 262–264.

A. R. Crathorne, *Moments de la binomiale par rapport à L'origine*, Comptes Rendus, vol. 198 (1934), p. 1202;

A. A. K. Aygangar, *Note on the recurrence formulae for the moments of the point binomial*, Biometrika, vol. 26 (1934), pp. 262–264.

To this, besides Soper's tract already mentioned, should be added:

Ch. Jordan, *Statistique Mathematique*, Paris, 1927.

K. Pearson, *On Certain Properties of the Hypergeometric Series* . . . , Phil. Mag., **47**, pp. 236–246 (1899).

2. **Moment Arrays.** As developed by Soper, frequency distributions may be exhibited by frequency arrays, in the case of a single variate, in the form:

$$f(A) = \sum_x p_x A^x \tag{2.1}$$

where p_x are the frequencies with which the measures, x, of the character, A, occur in a population.

The substitution $A = e^\alpha$ leads to the moment about the origin array:

$$\begin{aligned} f(e^\alpha) &= \sum_x p_x e^{x\alpha} \\ &= \sum_x p_x \left(1 + x\alpha + \frac{x^2\alpha^2}{2!} + \cdots\right) \\ &= \sum_s m_s \frac{\alpha^s}{s!} \end{aligned} \tag{2.2}$$

where

$$m_s = \sum_x p_x x^s$$

The symbol α is a logical or umbral symbol serving merely to identify the moments in the expansion of the array.

The moment array for moments about the mean is found from the relation:

$$\begin{aligned} \phi(e^\alpha) &= e^{-m_1\alpha} f(e^\alpha) \\ &= \sum_s \mu_s \alpha^s / s! \end{aligned}$$

where m_1 is the first moment about the origin.

The moment arrays for the distributions concerned are as follows:

Binomial $\quad f(e^\alpha) = [1 + p(e^\alpha - 1)]^n = \sum_{x=0}^{n} \binom{n}{x} p^x(e^\alpha - 1)^x$

Poisson $\quad f(e^\alpha) = e^{a(e^\alpha - 1)} = \sum_{x=0}^{\infty} \frac{a^x(e^\alpha - 1)^x}{x!}$

Hypergeometric $\quad f(e^\alpha) = \sum_{x=0}^{\infty} \frac{(l)_x(r)_x}{(n)_x} \frac{(e^\alpha - 1)^x}{x!}$

where the parameters p, n, and a for the binomial and Poisson have the usual significance. The parameters for the hypergeometric distribution, with the substitution $r = s$, follow Soper; Pearson (loc. cit.) uses q, r, n, where $q = l/n$. The notation $(l)_x$ means

$$(l)_x = l(l - 1) \cdots (l - x + 1).$$

It will be seen that, with the usual interpretation of $\binom{n}{x}$ as zero for $x > n$,

the three distributions so far as concerns α may be exhibited by a function of the form

$$f(e^\alpha) = \sum_{x=0}^{\infty} A_x(e^\alpha - 1)^x$$

where A_x of course depends on the distribution concerned.

3. **Moments About the Origin.** The moments about the origin can then be defined by the equation:

$$\sum_{s=0}^{\infty} m_s \frac{\alpha^s}{s\,!} = \sum_{x=0}^{\infty} A_x(e^\alpha - 1)^x \tag{3.1}$$

and

$$\sum_{x=0}^{\infty} A^x(e^\alpha - 1)^x = \sum_{x=0}^{\infty} A_x \sum_{v=0}^{x} (-1)^{x-v} \binom{x}{v} e^{v\alpha}$$
$$= \sum_{s=0}^{\infty} \frac{\alpha^s}{s\,!} \sum_{x=0}^{s} x\,!\, A_x S_{x,s}$$

where $S_{x,s}$ is a Stirling number of the second kind, as used by Jordan (loc. cit.) and defined by

$$x\,!\, S_{x,s} = \sum_{v=0}^{x} (-1)^{x-v} \binom{x}{v} v^s = \Delta^x\, 0^s,$$

$\Delta^x 0^s$ being in the language of the finite difference calculus, a "difference of nothing" that is $\Delta^x n^s \mid n = 0$.

The internal series terminates at s because $S_{x,s} = 0$, $x > s$, as is readily apparent in the finite difference expression. Further $S_{0,s} = 0$, $s \neq 0$; $S_{0,0} = 1$.

By equating coefficients in equation (3.1), m_s, the sth moment about the origin, is given by

$$m_s = \sum_{x=0}^{s} x\,!\, A_x\, S_{x,s}. \tag{3.2}$$

The particular forms for the three distributions are as follows:

$$m_s = \sum_{x=0}^{s} (n)_x\, p^x\, S_{x,s} \qquad \textit{Binomial} \tag{3.3}$$

$$m_s = \sum_{x=0}^{s} a^x\, S_{x,s} \qquad \textit{Poisson} \tag{3.4}$$

$$m_s = \sum_{x=0}^{s} \frac{(l)_x (r)_x}{(n)_x}\, S_{x,s} \qquad \textit{Hypergeometric} \tag{3.5}$$

The Stirling numbers have the following recurrence relation (Jordan loc. cit.):

$$S_{x,s+1} = xS_{x,s} + S_{x-1,s}. \tag{3.6}$$

This relation in conjunction with equations (3.3)–(3.5) leads to moment recurrence relations. The procedure is illustrated for the binomial distribution as follows:

$$\begin{aligned} m_{s+1} &= \sum_{x=0}^{s+1} (n)_x\, p^x\, S_{x,\,s+1} \\ &= \sum_{x=0}^{s+1} (n)_x\, p^x\, (x\, S_{x,\,s} + S_{x-1,\,s}) \\ &= p\, D_p\, m_s + (npm_s - p^2\, D_p\, m_s) \\ &= npm_s + pq\, D_p\, m_s \end{aligned}$$

where $q = 1 - p$.

The steps in the process are expanded as follows:

$$\begin{aligned} \sum_{x=0}^{s+1} (n)_x p^x\, x\, S_{x,\,s} &= \sum_{x=0}^{s} (n)_x\, p^x\, x\, S_{x,\,s} \\ &= \sum_{x=0}^{s} (n)_x\, S_{x,\,s}\, pD_p(p^x) \\ &= pD_p\, m_s \end{aligned}$$

$$\begin{aligned} \sum_{x=0}^{s+1} (n)_x\, p^x\, S_{x-1,\,s} &= \sum_{x=0}^{s+1} (n - x + 1)\, (n)_{x-1}\, p^x\, S_{x-1,\,s} \\ &= n \sum_{x=1}^{s} (n)_x\, p^{x+1}\, S_{x,\,s} - \sum_{x=1}^{s} x(n)_x\, p^{x+1}\, S_{x,\,s} \\ &= np\, m_s - p^2 D_p\, m_s \end{aligned}$$

The results for the three distributions are as follows:

(3.7) $\qquad m_{s+1} = npm_s + pqD_pm_s \qquad$ *Binomial*

(3.8) $\qquad m_{s+1} = am_s + aD_am_s \qquad$ *Poisson*

(3.9) $\qquad m_{s+1} = \dfrac{lr}{n}\, m_s(l - 1, r - 1, n - 1) - (n + 1)\Delta_n m_s \qquad$ *Hypergeometric*

Here D_p and D_a denote differentiation with respect to p and a, respectively, and Δ_n denotes the difference operation with respect to n. For the hypergeometric distribution the moments are functions of l, r, and n as well as of s; $m_s(l - 1, r - 1, n - 1)$ is the same function of $l - 1$, $r - 1$ and $n - 1$ as $m_s(l, r, n)$ is of l, r, n. Equation (3.9) appears to be new.

For convenience of reference, a short table of the Stirling numbers of the second kind follows:

$s \setminus x$	$S_{x,s}$ 0	1	2	3	4	5
0	1					
1	0	1				
2	0	1	1			
3	0	1	3	1		
4	0	1	7	6	1	
5	0	1	15	25	10	1

4. **Moments About the Mean.** As shown in Section 2 above, moments about the mean may be defined as follows:

$$\sum_{s=0}^{\infty} \mu_s \frac{\alpha^s}{s!} = \sum_{x=0}^{\infty} A_x e^{-m_1\alpha} (e^\alpha - 1)^x \tag{4.1}$$

where m_1 is the first moment about the origin:

$$\begin{aligned} m_1 &= np \quad \textit{Binomial} \\ &= a \quad \textit{Poisson} \\ &= lr/n \; \textit{Hypergeometric} \end{aligned}$$

Now

$$\begin{aligned} \sum_{x=0}^{\infty} A_x e^{-m_1\alpha} (e^\alpha - 1)^x &= \sum_{x=0}^{\infty} A_x \sum_{v=0}^{x} (-1)^{x-v} \binom{x}{v} e^{(v-m_1)\alpha} \\ &= \sum_{s=0}^{\infty} \frac{\alpha^s}{s!} \sum_{x=0}^{s} x!\, A_x \sigma_{x,s} \end{aligned}$$

where

$$x!\, \sigma_{x,s} = \sum_{v=0}^{x} (-1)^{x-v} \binom{x}{v} (v - m_1)^s = \Delta^x (-m_1)^s.$$

It will be observed that for $m_1 = 0$, $\sigma_{x,s} = S_{x,s}$. The internal series terminates at s for the same reason as before.

The moments about the mean are then given by:

$$\mu_s = \sum_{x=0}^{s} x!\, A_x \sigma_{x,s} \tag{4.2}$$

The particular forms for the three distributions are as follows:

$$\mu_s = \sum_{x=0}^{s} (n)_x p_x \sigma_{x,s} \qquad \textit{Binomial} \tag{4.3}$$

$$\mu_s = \sum_{x=0}^{s} a^x \sigma_{x,s} \qquad \textit{Poisson} \tag{4.4}$$

$$\mu_s = \sum_{x=0}^{s} \frac{(l)_x (r)_x}{(n)_x} \sigma_{x,s} \qquad \textit{Hypergeometric.} \tag{4.5}$$

The coefficients $\sigma_{x,s}$ satisfy the following recurrence relation:[4]

$$\sigma_{x,s+1} = (x - m_1)\sigma_{x,s} + \sigma_{x-1,s} \tag{4.6}$$

which in conjunction with equations (4.3)–(4.5) leads to moment recurrence relations as before. The actual derivation is somewhat complicated by the circumstance that $\sigma_{x,s}$ is a function of m_1 and therefore of the frequency parameters, rather than a constant as before. The derivation is illustrated for the binomial distribution as follows:

$$\begin{aligned}
\mu_{s+1} &= \sum_{x=0}^{s+1} (n)_x\, p^x \sigma_{x,\,s+1} \\
&= \sum_{x=0}^{s+1} (n)_x\, p^x [(x - np)\sigma_{x,\,s} + \sigma_{x-1,\,s}] \\
&= \sum_{x=0}^{s} (n)_x \sigma_{x,\,s}\, pD_p(p^x) - np\mu_s + \sum_{x=0}^{s+1} (n)_x\, p^x \sigma_{x-1,\,s} \\
&= pD_p\mu_s + nsp\mu_{s-1} - np\mu_s + np\mu_s - p^2[D_p\mu_s + ns\mu_{s-1}] \\
&= pq\,[ns\mu_{s-1} + D_p\mu_s].
\end{aligned}$$

The steps in the process are expanded as follows:

$$\begin{aligned}
\sum_{x=0}^{s} (n)_x \sigma_{x,\,s}\, pD_p(p^x) &= \sum_{x=0}^{s} (n)_x [pD_p(p^x \sigma_{x,\,s}) - p^x\, pD_p(\sigma_{x,\,s})] \\
&= pD_p\mu_s - p \sum_{x=0}^{s} (n)_x\, p^x(-\, ns\sigma_{x,\,s-1}) \\
&= pD_p\mu_s + nsp\mu_{s-1} \\
\sum_{x=0}^{s+1} (n)_x p^x \sigma_{x-1,\,s} &= \sum_{x=0}^{s+1} (n - x + 1)\,(n)_{x-1}\, p^x \sigma_{x-1,\,s} \\
&= n \sum_{x=0}^{s} (n)_x\, p^{x+1} \sigma_{x,\,s} - \sum_{x=0}^{s} x(n)_x\, p^{x+1} \sigma_{x,\,s} \\
&= np\,\mu_s - p^2\,[D_p\,\mu_s + ns\,\mu_{s-1}].
\end{aligned}$$

The relation $D_p\sigma_{x,\,s} = -ns\sigma_{x,\,s-1}$ is obtained from the definition equation of $\sigma_{x,\,s}$ (with $m_1 = np$).

The resulting recurrence relations for the three distributions are as follows:

$$\mu_{s+1} = nspq\,\mu_{s-1} + pq\,D_p\,\mu_s \qquad \textit{Binomial} \tag{4.7}$$

$$\mu_{s+1} = as\mu_{s-1} + a\,D_a\,\mu_s \qquad \textit{Poisson} \tag{4.8}$$

---

[4] Jordan, loc. cit. or E. C. Molina, *An Expansion for Laplacian Integrals . . .*, Bell System Technical Journal, **11**, p. 571.

$$\mu_{s+1} = (n+1)\left[\mu_s - \sum_{v=0}^{s} \binom{s}{v} K_1^v \mu_{s-v}(l, r, n+1)\right] \quad \textit{Hypergeometric}$$
$$- \frac{lr}{n}\left[\mu_s - \sum_{v=0}^{s} \binom{s}{v} K_2^v \mu_{s-v}(l-1, r-1, n-1)\right] \tag{4.9}$$

where

$$K_1 = \frac{-lr}{n(n+1)} = \Delta_n \frac{lr}{n}$$
$$K_2 = \frac{(l-1)(r-1)}{(n-1)} - \frac{lr}{n}.$$

The last of these, which appears to be new, seems to be of formal interest only.

The coefficients $\sigma_{x,s}$ are related to the Stirling numbers by the expression:

$$\sigma_{x,s} = \sum_{v=0}^{s-x} (-1)^v \binom{s}{v} S_{x,s-v} m_1^v = \sum_{v=0}^{s-x} a_v m_1^v$$

and consequently can be exhibited with detached coefficients in the form $a_0 + a_1 + a_2 + \cdots + a_{s-x}$. For the binomial and Poisson distributions certain simplifications, to be developed in the section following, in equations (4.3) and (4.4) may be made. For the hypergeometric distribution it appears necessary to use equation (4.5); the following short table of $\sigma_{x,s}$, employing the detached coefficients mentioned above, is given for this purpose:

$s \backslash x$	$\sigma_{x,s}$: 0	1	2	3	4	5
1	0−1	1				
2	0+0+1	1−2	1			
3	0+0+0−1	1−3+3	3−3	1		
4	0+0+0+0+1	1−4+6−4	7−12+6	6−4	1	
5	0+0+0+0+0−1	1−5+10−10+5	15−35+30−10	25−30+10	10−5	1

5. Binomial and Poisson Moments About the Mean—Simplified Formulas.

5.1 Binomial. From examination of the first few moments about the mean, it appears expedient[5] to write the formulas:

$$\mu_{2s} = \sum_{x=1}^{s} \alpha_{x,2s} (npq)^x$$
$$\mu_{2s+1} = (q-p) \sum_{x=1}^{s} \alpha_{x,2s+1} (npq)^x \tag{5.1.1}$$

---

[5] The kind of expression chosen admits of some variety. A recurrence relation for coefficients in the expansion $\mu_s = \sum_{x=1}^{s} \alpha_{x,s}\, p^x$ has been given by E. H. Larguier, *On a Method For Evaluating the Moments of a Bernoulli Distribution*, Bull. Am. Math. Soc., **42**, 1, p. 24 (Abstract 8); I am indebted to Mr. Larguier for the opportunity of examining his results in advance of publication.

When these are substituted into the moment recurrence relation, the coefficients are found to be related as follows:

$$\alpha_{x,2s} = [x + pqD_{pq}]\alpha_{x,2s-1} + (2s-1)\alpha_{x-1,2s-2} - 2pq[1 + 2x + 2pqD_{pq}]\alpha_{x,2s-1}$$

$$\alpha_{x,2s+1} = [x + pqD_{pq}]\alpha_{x,2s} + 2s\alpha_{x-1,2s-2}$$

or, in general,

$$\alpha_{x,s+1} = [x + pqD_{pq}]\alpha_{x,s} + s\alpha_{x-1,s-1} - pq[1-(-1)^s]\,[1 + 2x + 2pqD_{pq}]\alpha_{x,s} \tag{5.1.2}$$

Using detached coefficients of powers of pq as outlined above, these coefficients may be exhibited as follows:

$s \backslash x$	$\alpha_{x,s}$: 1	2	3	4
2	1			
3	1			
4	1 − 6	3		
5	1 − 12	10		
6	1 − 30 + 120	25 − 130	15	
7	1 − 60 + 360	56 − 462	105	
8	1 − 126 + 1680 − 5040	119 − 2156 + 7308	490 − 2380	105
9	1 − 252 + 5040 − 20160	246 − 6948 + 32112	1918 − 13216	1260

It may be noted that the coefficients of the first column in conjunction with equations (5.1.1) give the binomial seminvariants.

Equations (5.1.1) make the coefficients functions of pq only; a slight alteration makes the coefficients functions of n only. Thus:

$$\mu_{2s} = \sum_{x=1}^{s} \beta_{x,2s}(pq)^x \tag{5.1.3}$$

$$\mu_{2s+1} = (q-p)\sum_{x=1}^{s} \beta_{x,2s+1}(pq)^x$$

and the coefficients are found to satisfy the recurrence relation:

$$\beta_{x,s+1} = x\beta_{x,s} + ns\beta_{x-1,s-1} - [1-(-1)^s](2x-1)\beta_{x-1,s}. \tag{5.1.4}$$

These coefficients may be exhibited by a rearrangement of the table given

above as may be seen by comparing equations (5.1.1) and (5.1.3). The first few coefficients are as follows:

$s \setminus x$	$n^{-1}\beta_{x,s}$		
	1	2	3
2	1		
3	1		
4	1	$-6+3$	
5	1	$-12+10$	
6	1	$-30+25$	$120-130+15$

5.2 **Poisson.** The Poisson moments about the mean may be expressed as follows:

$$\mu_s = \sum_{x=0}^{[s/2]} \alpha_{x,s}\,\alpha^x \tag{5.2.1}$$

where [ ] represents "integral part of" and

$$\alpha_{x,s+1} = x\alpha_{x,s} + s\alpha_{x-1,s-1}. \tag{5.2.2}$$

The coefficients $\alpha_{x,s}$ are the constant terms in the expressions for the corresponding binomial distribution coefficients in powers of pq.

BELL TELEPHONE LABORATORIES.

NOTE ON ZOCH'S PAPER ON THE POSTULATE OF THE ARITHMETIC MEAN

By Albert Wertheimer

1. **Introduction.** There appeared recently a paper by Richmond T. Zoch[1] entitled "On The Postulate of the Arithmetic Mean." The stated purpose of his paper, was to show that the derivation of the Postulate as given by Whittaker & Robinson, is not correct. It is the purpose of this paper to show, that Zoch has not proven any error to exist in the Whittaker & Robinson derivation, but that there are a few errors in his paper. As this paper is intended to be read with Zoch's paper as a reference, the terms used there will not be redefined here, and except where otherwise stated, the symbols used will have the same meaning.

2. Zoch introduces the function

$$f \equiv \bar{x} + a\mu_3/\mu_2$$

and claims that it satisfies all the four axioms of Whittaker & Robinson, and obviously it is not the arithmetic mean. He therefore concludes that their derivation must have errors somewhere, and proceeds to find them. Let us first examine the f function. Considering only the part μ_3/μ_2, the partial derivatives with respect to x_i are given by

$$\frac{3\mu_2\{(x_i - \bar{x})^2 - \mu_2\} - 2\mu_3(x_i - \bar{x})}{n\mu_2^2}$$

It is then stated (p. 172) "... clearly these partial derivatives are single valued and continuous. Therefore the function μ_3/μ_2 satisfies axiom IV." Now, the condition that a function be continuous and single valued means of course that this be true throughout the region of definition of the function. It is not shown how these derivatives are clearly continuous and single valued for the very important case where all the x's are equal and the derivatives become indeterminate. As a matter of fact they are not continuous in this case, and therefore the f function does not satisfy axiom IV. To prove this, we only have to consider the very simple case where we let

$$x_i = k + c_i z$$

---

[1] This Journal Vol. VI no. 4, Dec. 1935, pp. 171–182.

where k is a fixed constant, c_i is a set of arbitrary constants not all equal, and z is a parameter. We then have

$$\bar{x} = k + \bar{c}z$$
$$\mu_2 = \mu_2' z^2$$
$$\mu_3 = \mu_3' z^3$$

where

$$\bar{c} = 1/n \sum c_i$$
$$\mu_2' = 1/n \sum (c_i - \bar{c})^2$$
$$\mu_3' = 1/n \sum (c_i - \bar{c})^3$$

Substituting these values in f and the derivatives, we get taking $a = 1$,

$$f = k + z\bar{c} + z^3\mu_3'/z^2\mu_2'$$

$$\partial f/\partial x_i = 1/n + \frac{3z^2\mu_2'\{z^2(c_i - \bar{c})^2 - z^2\mu_2'\} - 2z^4\mu_3'(c_i - \bar{c})}{nz^4\mu_2'^2}$$

Now going to the limit when z approaches zero, and all the x's approach k, we get

$$\lim_{z \to 0} f = k,$$

$$\lim_{z \to 0} \partial f/\partial x_i = 1/n\{-2 + 3(c_i - \bar{c})^2/\mu_2' - 2\,\mu_3'(c_i - \bar{c})/\mu_2'^2\}$$

Thus, when all the x's approach the same value, the function f also approaches the same value independent of the c's, that is regardless of the mode of approach, while the derivatives can take on any value depending on the c's that is on how the limiting value of f is approached. The f function then does not have continuous single valued partial derivatives, and therefore does not satisfy axiom IV.

In part 2 of the paper it is stated "Now when the x_i all approach a then both f and $\partial f/\partial x_j$ become indeterminate forms. However, in this case f takes an indeterminate form which can be evaluated and it can be shown that μ_3/μ_2 will always have the value zero, i.e., f will have the value a when all the $x_i \to a$; while the $\partial f/\partial x_i$ can take any value whatever and in general the $\partial f/\partial x_i$ will not be equal when the $x_i \to a$." This statement really amounts to saying that the f function does not satisfy axiom IV, but it is there used to demonstrate that one of Schiaparelli's propositions is false.

3. Having exhibited a function different from the arithmetic mean, and supposedly satisfying all the four axioms, the question is asked "Where is the proof given by Whittaker & Robinson lacking in rigor?" After numbering the various steps in the derivation ". . . for the sake of rigor and careful reasoning

. . ." it is stated (p. 174), "The sixth step involves the tacit assumption that the partial derivatives are functions of k. These partial derivatives are not necessarily functions of k . . ." and it is therefore concluded that the sixth step is not valid. Now, how can any function that by definition is to be evaluated at $\theta k x_i$ not be a function of k? What is shown (pp. 174–5) is that these derivatives do not necessarily involve k explicitly, but this is neither implied nor necessary for the sixth step, and there is no ground for doubting its validity.

4. In order to overcome the supposed defect in the sixth step, it is proposed to change axiom IV so as to require the partial derivatives to be constants. But even then (p. 175) ". . . there remains an objection in the seventh step." Now, the seventh step consists of the statement that if

$$\phi(x_i) = \sum c_i x_i$$

where the c's are independent of the x's then due to the condition that ϕ be a symmetric function, all the c's must be equal. To show the defect in this step it is stated, that under certain conditions ". . . the function $f \equiv \bar{x} + \mu_3/\mu_2$ will have partial derivatives with respect to x_i which are unequal and constant; yet at the same time the function f is a symmetrical expression of the n variables." Granting that all that is correct, what has this got to do with the seventh step? The f function certainly is not of the type $\sum c_i x_i$ to which the seventh step is applied.

5. One more point should be mentioned. On p. 181 it is supposedly proven that any function satisfying the first three axioms must have continuous first partial derivatives. The proof is essentially as follows: Assuming all the x's are given the same increment Δx, the increment of the function then is $\Delta\phi$. It is then stated ". . . but by axiom I, $\Delta\phi = \Delta x$. Therefore $\Delta\phi/\Delta x = 1 = d\phi/dx$. In other words, the total derivative of ϕ exists and is constant. Therefore the total derivative of ϕ is continuous." From this, the continuity of the first partial derivatives is proven by means of Euler's Theorem for homogeneous functions. Now, just what does the symbol $d\phi/dx$ (which is called the total derivative) mean for a function of many independent variables? Besides, (whatever this symbol means) is it considered rigorous to deduce a general Theorem from the very special case where all the differentials are made equal? This is one place where the f function could be used effectively as an exhibit of a function satisfying the first three axioms, and not having continuous partial derivatives.

It is also stated (p. 181) that ". . . it would seem more satisfactory to postulate that the function ϕ is single valued, for the single-valuedness of a derivative does not insure the single-valuedness of the integral while the single-valuedness of a function does insure the single-valuedness of the derivative where the derivative exists." This statement is certainly not self evident and requires

proof. For a single variable at least, it is easy to imagine a function represented by a curve with corners defined in a certain interval. The function then could be single valued everywhere in the interval, while the derivatives at the corners may exist and have two distinct values, depending on whether the corner is approached from the right or the left. On the other hand it is hard to imagine a curve representing a single valued function such that the integral i.e. the function represented by the area under the curve should not be single valued.

6. **In Conclusion:** It is stated in the Introduction that "Since this book has had wide circulation, it is believed that the errors in this proof should be called to the attention of the users of the book. The present paper has been prepared for this purpose." It is for the same reason, that this paper was prepared to show that no error has been proven to exist.

BUREAU OF ORDNANCE, U. S. NAVY DEPARTMENT

NOTE ON THE BINOMIAL DISTRIBUTION

By C. E. Clark

The purpose of this note is to show that

$$f(x) = (-1)^n \frac{q^n n!}{\pi} \left(\frac{p}{q}\right)^x \frac{\sin \pi x}{x^{(n+1)}} \tag{1}$$

where n is an integer $\geqq 0$, $0 < p < 1$, $p + q = 1$, and $x^{(n+1)} = x(x-1)(x-2) \cdots (x-n)$, is a function whose values at $x = 0, 1, 2, \cdots n$ are the successive terms of the expansion of $(q + p)^n$, and also to consider the problem of fitting $f(x)$ to an observed frequency distribution.

The statement made about (1) can be verified by evaluating (1) as an indeterminate form. On the other hand, (1) can be derived by observing that the x-th term (x an integer) of the expansion of $(q + p)^n$ is

$$\frac{n!}{x!(n-x)!} p^x q^{n-x} = \frac{\Gamma(n+1) p^x q^{n-x}}{\Gamma(x+1)\Gamma(n-x+1)}; \tag{2}$$

then (1) can be derived from (2) by means of the product expansions for $\Gamma(x)$ and $\sin x$. This derivation of (1) from (2) can also be carried out by expressing (2) as a Beta function and then using

$$B(x+1, n-x+1) = \int_0^\infty \frac{t^x}{(1+t)^{n+2}} dt = (-1)^n \frac{\pi}{(n+1)!} \frac{x^{(n+1)}}{\sin \pi x}.$$

This integration can be performed by means of the theory of residues.

Consider the problem of fitting (1) to an observed frequency distribution. We shall write (1) in the form

$$F(z) = ab^x \frac{\sin \pi x}{x^{(n+1)}}, \qquad x = \frac{nb}{1+b} + h(z - \bar{z}) \tag{3}$$

and determine the constants a, b, n, and h so that, when $\bar{z}$ is the mean of the observed distribution, $F(z)$ will fit the distribution.

The values of a, b, n, and h can be determined by the method of moments. Let ν_2, ν_3, and ν_4, denote the usual second, third, and fourth moments of the distribution, which are calculated in the usual way (as in W. P. Elderton, *Frequency-Curves and Correlation*) and not adjusted by any procedure such as Sheppard's adjustments. Also, use the usual notation $\beta_1 = \frac{\nu_3^2}{\nu_2^3}$ and $\beta_2 = \frac{\nu_4}{\nu_2^2}$.

Then, the method of moments gives

$$n = \frac{2}{3 + \beta_1 - \beta_2} \tag{4}$$

$$b = \frac{2 + n\beta_1 \pm \sqrt{n\beta_1(4 + n\beta_1)}}{2} \tag{5}$$

$$h = \sqrt{\frac{nb}{\nu_2}}\left(\frac{1}{1+b}\right)$$

$a = (-1)^n \dfrac{h(\Sigma f)n!}{\pi(1+b)^n}$, where Σf is the sum of the frequencies of the distribution. An integer n is chosen nearest the value assigned by (4). The two values of b from (5) determine two curves that are congruent but whose skewnesses are of opposite sign. Hence, b is uniquely determined by (5) and the sign of the skewness of the data.

For a symmetrical distribution, $b = 1$, $\nu_3 = 0$, and

$$n = \frac{2}{3 - \beta_2}$$

$$h = \frac{\sqrt{n}}{2\sqrt{\nu_2}}.$$

We shall consider an illustrative example. In the following table the columns $f(z)$ and $f_2(z)$ are taken from W. P. Elderton, *Frequency-Curves and Correlation* (1906), page 62. $f(z)$ is an empirical frequency distribution, while $f_2(z)$ is obtained by fitting a Pearson Type II curve to the distribution $f(z)$. $f_1(z)$ is computed from

$$f_1(z) = 1624\,\frac{\sin \pi x}{x^{(6)}}, \qquad x = 2.0973 + .808z$$

which is determined by the method of this note. $f_3(z)$ is obtained by fitting the normal curve

$$f_3(z) = 485.1e^{-\frac{(z-.4985)^2}{2(1.829)}}.$$

z	$f(z)$	$f_1(z)$	$f_2(z)$	$f_3(z)$
−3	11	18	14	19
−2	116	107	109	92
−1	274	281	286	263
0	451	438	433	444
1	432	437	433	444
2	267	267	285	263
3	116	106	109	92
4	16	18	14	19

The coefficients of goodness of fit for $f_1(z)$, $f_2(z)$, and $f_3(z)$ are respectively .35, .58, and .02.

CONVEXITY PROPERTIES OF GENERALIZED MEAN VALUE FUNCTIONS[1]

By Nilan Norris

Consider the following generalized mean value functions: (1) the unit weight or simple sample form, $\phi(t) = \left(\frac{x_1^t + x_2^t + \cdots + x_n^t}{n}\right)^{\frac{1}{t}}$, in which the x_i are positive real numbers not all equal each to each, and in which t may take any real value; (2) the weighted sample form, $\omega(t) = \left(\frac{c_1x_1^t + c_2x_2^t + \cdots + c_nx_n^t}{c_1 + c_2 + \cdots + c_n}\right)^{\frac{1}{t}}$, in which the c_i are positive numbers not all equal each to each, and in which the x_i and t are restricted as in $\phi(t)$; (3) the integral form, $\theta(t) = \left[\int_{x=0}^1 x^t dx\right]^{\frac{1}{t}}$, where $\int_{x=0}^1 x^t dx$ exists for every real value of t; and (4) the generalized integral form $\Phi(t) = \left[\int_{x=0}^\infty x^t d\psi(x)\right]^{\frac{1}{t}}$, where $\psi(x)$ is a non-decreasing function integrable in the Riemann-Stieltjes sense such that $\psi(\infty) - \psi(0) = 1$, and such that $\int_{x=0}^\infty x^t d\psi(x)$ exists for every real value of t. The facts that all of these functions are monotonic increasing and that both $\phi(t)$ and $\omega(t)$ have two horizontal asymptotes have been previously demonstrated.[2] Although the existence of $\phi(t)$ and $\omega(t)$ has been known since 1840, there appears to have been no attempt made to investigate the behavior of the second derivatives of them.[3]

When the x_i are price relatives, production relatives, or similar data, $\phi(t)$ and $\omega(t)$ yield common types of index numbers by direct substitution of integral values of t. For any values of t such that $0 < t_1 < t_2 < \infty$, the type bias of $\phi(t_2)$ will be greater than the type bias of $\phi(t_1)$. Similarly, for any values of t such that $-\infty < t_1 < t_2 < 0$, the type bias of $\phi(t_1)$ will be greater than the type bias of $\phi(t_2)$. The second derivatives of $\phi(t)$ and $\omega(t)$ indicate whether

---

[1] Presented at a joint meeting of the American Mathematical Society, the Econometric Society, and the Institute of Mathematical Statistics at St. Louis on January 2, 1936. The writer is indebted to C. C. Craig, Einar Hille, Dunham Jackson, and J. Shohat for helpful critical reviews of the preliminary draft of this paper.

[2] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities* (Cambridge University Press, London, 1934), pp. 12–15; and Nilan Norris, "Inequalities among Averages," *Annals of Mathematical Statistics*, Vol. VI, No. 1, March, 1935, pp. 27–29.

[3] Jules Bienaymé, *Société Philomatique de Paris*, Extraits des procès-verbaux des séances pendant l'année 1840 (Imprimerie D'A. René et Cie., Paris, 1841), Séance du 13 juin 1840 p. 68.

type bias is changing at an increasing or a decreasing rate as between the unlimited number of averages available for use. Considerable interest attaches to $\omega(t)$, the weighted sample form of function.

Let $\omega(t)$ be made arbitrary for the case of $n = 2$, with $x_1 = 1$, and $x_2 = e^{-\lambda}$, where λ is any real number. Also let $c_1 = \alpha$, and $c_2 = \beta$, where $\alpha + \beta = 1$. Then $\omega(t) = [\alpha + \beta e^{-\lambda t}]^{\frac{1}{t}}$. Now for all values of t,

$$\alpha + \beta e^{-\lambda t} = 1 - \frac{\beta\lambda}{1} t + \frac{\beta\lambda^2}{2} t^2 - \frac{\beta\lambda^3}{6} t^3 + \cdots$$

For $|t|$ sufficiently small, it follows that

$$\log(\alpha + \beta e^{-\lambda t}) = -\beta\lambda t + \tfrac{1}{2}\beta\lambda^2(1-\beta)t^2 + \beta\lambda^3\left[-\frac{1}{6} + \frac{\beta}{2} - \frac{\beta^2}{3}\right]t^3 + \cdots,$$

so that for $t \neq 0$

$$\frac{1}{t}\log(\alpha + \beta e^{-\lambda t}) = -\beta\lambda + \tfrac{1}{2}\beta\lambda^2(1-\beta)t + \beta\lambda^3\left[-\frac{1}{6} + \frac{\beta}{2} - \frac{\beta^2}{3}\right]t^2 + \cdots.$$

Therefore $\omega(t) = \exp.\left[\frac{1}{t}\log(\alpha + \beta e^{-\lambda t})\right]$

$$= e^{-\beta\lambda}\left[1 + \tfrac{1}{2}\beta\lambda^2(1-\beta)t + \beta\lambda^3\left\{-\frac{1}{6} + \frac{\beta}{2} - \frac{\beta^2}{3} + \frac{1}{8}\beta(1-\beta)^2\lambda\right\}t^2 + \cdots\right].$$

It follows that $\omega''(0) = 2\beta\lambda^3 e^{-\beta\lambda}\left[-\frac{1}{6} + \frac{\beta}{2} - \frac{\beta^2}{3} + \frac{1}{8}\beta(1-\beta)^2\lambda\right]$. It is clear that $\omega(0)$ is the weighted geometric mean, and that $\phi(0)$ is the unit weight or simple sample form of geometric mean. As a means of demonstrating the range of values which $\omega''(0)$ may take it is helpful to rewrite the expression for $\omega''(0)$ as follows:

$$\omega''(0) = \tfrac{1}{4}\beta^2(1-\beta)^2\lambda^3\left[\lambda - \frac{4}{3}\frac{1-2\beta}{\beta(1-\beta)}\right]e^{-\beta\lambda} \equiv f(\lambda, \beta).$$

This consideration makes it possible to distinguish three cases of $y = f(\lambda, \beta)$ for fixed β, namely, $0 < \beta < \frac{1}{2}$; $\beta = \frac{1}{2}$; and $\frac{1}{2} < \beta < 1$. In all three cases $f(\lambda, \beta)$ has an absolute minimum $\mu(\beta) \leqq 0$, and $\mu(\frac{1}{2}) = 0$. The corresponding values of λ satisfies the quadratic equation $\lambda^2 - \frac{4}{3}\frac{4-5\beta}{\beta(1-\beta)}\lambda + \frac{4-8\beta}{\beta^2(1-\beta)} = 0$. It is clear that by taking β near enough to 0, one can make $\mu(\beta)$ as large negative as is desired. Also, by choosing λ properly, one can make $\omega''(0)$ take any value between $\mu(\beta)$ and ∞. For example, when $\alpha = \beta = \frac{1}{2}$, λ may be selected so as to make $\omega''(0)$ any arbitrarily chosen non-negative number. For then $\omega''(0) = \frac{\lambda^4}{64}e^{-\frac{\lambda}{2}}$, and as λ increases from $-\infty$ to 0, $\omega''(0)$ decreases from ∞ to 0. If $\lambda = 0, \omega''(0) = 0$. If $\lambda > 0$, as λ increases from 0 to 8, $\omega''(0)$ increases to

$64e^{-4}$, and as λ increases beyond 8, $\omega''(0)$ decreases, approaching 0 as λ increases indefinitely. It is evident that the case of $\alpha = \beta = \frac{1}{2}$, with $\lambda = -\log 2$, $x_1 = 1$, and $x_2 = e^{-\lambda}$, is one in which $\omega(t)$ becomes the unit weight or simple sample type of generalized mean value function, namely, $\phi(t) = \left(\frac{1^t + 2^t}{2}\right)^{\frac{1}{t}}$. Reference to the first expression above noted for $\omega''(0)$ will make clear that $\phi''(0) = \frac{(\log 2)^4}{64}\sqrt{2}$ in this special case.

Analysis of $\Phi(t)$, the generalized integral form of generalized mean value function, makes it possible to characterize populations of a very general character, as well as samples. But in the case of $\Phi(t)$ it is even more difficult to generalize as to convexity properties. For example, let

$$\Phi(t) = \left[\int_{u=-\infty}^{\infty} e^{-ut}\, dE(u)\right]^{\frac{1}{t}},$$

where

$$E(u) = \frac{1}{\sqrt{\pi}} \int_{v=-\infty}^{u} e^{-v^2}\, dv.$$

This expression is obviously of the required generalized integral type. Now

$$[\Phi(t)]^t = \frac{1}{\sqrt{\pi}} \int_{u=-\infty}^{\infty} e^{-ut-u^2}\, du = \frac{1}{\sqrt{\pi}}\, e^{\frac{t^2}{4}} \int_{u=-\infty}^{\infty} e^{-\left(u+\frac{t}{2}\right)^2}\, du = e^{\frac{t^2}{4}}.$$

Therefore $\Phi(t) = e^{\frac{t}{4}}$, and $\Phi''(t) = \dfrac{e^{\frac{t}{4}}}{16} > 0$ for all t. That is, in this particular case, $\Phi(t)$ has only one horizontal asymptote.

The foregoing examples indicate that the following conclusions may be drawn as to the diverse convexity attributes of the various means as functions of t: (1) The unit weight form, $\phi(t)$, and the weighted sample form, $\omega(t)$, must always have a point of inflection, since both of them not only increase with t, but are doubly asymptotic (have two horizontal asymptotes). (2) Points of inflection for $\phi(t)$ and $\omega(t)$ do not necessarily occur at $t = 0$. (3) The generalized integral form, $\Phi(t)$, need not always have a point of inflection. That is, the second derivatives of certain forms of $\Phi(t)$ do not change their sign, since such forms are concave upward.

UNIVERSITY OF MICHIGAN.

A SIMPLE FORM OF PERIODOGRAM

By Dinsmore Alter

Schuster's introduction of a method of systematic search for hidden periodicities and cycles opened a new field for the investigator of statistical data. The beauty of his method in its analogy to analysis of light, and the great reputation of its author, combined to give it universal acceptance and to blind statisticians to its faults.

In more recent years at least three new mathematical and two mechanical forms of periodogram analysis have been proposed, each of which exhibits certain advantages over the original one. The use of the term *periodogram* for these forms is an extension of Schuster's original definition which used as abscissae quantities proportional to the squares of the amplitudes of the sine terms found in the data for the various trial periods. He wrote: "It is convenient to have a word for some representation of a variable quantity which shall correspond to the spectrum of a luminous radiation. I propose the word *periodogram* and define it more particularly in the following way:

Let $\frac{1}{2}Ta = \int_{t_1}^{t_1+T} f(t) \cos ktdt$ and $\frac{1}{2}Tb = \int_{t_1}^{t_1+T} f(t) \sin ktdt$

where T may for convenience be chosen equal to some integer multiple of $\frac{2\pi}{k}$, and plot a curve with $\frac{2\pi}{k}$ as abscissae and $r = \sqrt{a^2 + b^2}$ as ordinates; this curve, or better, the space between this curve and the axis of abscissae, represents the periodogram of $f(t)$."

The following appear to be the essential criteria for a satisfactory form of periodogram:

1. It must exhibit plainly any repetition of form in the data regardless of how irregular the shape of the repeated interval may be. In doing this it must exaggerate the amplitude of the main terms at the expense of the lesser ones.
2. The calculation of the indices must be short. In a periodogram from many data the indices sometimes are computed for several hundred trial periods.
3. There should be a geometrical interpretation of the index used.
4. The frequency distribution of the index must be known.
5. Combining or smoothing the data should modify the index in a manner which leaves an obvious interpretation.

The Schuster periodogram has the following disadvantages:

1. Only sine terms of large amplitude are exhibited. A perfect repetition of an extremely irregular form of data would not be indicated in any way.
2. The calculations are long.
3. There is a considerable uncertainty in the length of the period found. Those methods of analysis which use harmonics as well as the fundamental have much less of this uncertainty.

The correlation periodogram has advantages in each of these points over the Schuster. However, even with it the calculations are fairly long. Furthermore, the modification of the coefficient introduced by grouping or smoothing is not a linear one.

The periodogram described here is a slight modification of one for which a preliminary note was published in 1933. Additional features have been studied and its applications to many data have shown its ease of calculation. This calculation has been reduced still more by a mechanical method which renders it practicable to contemplate the possibility of studying many data hitherto prohibited by excessive cost.

Consider data $x_0, x_1, x_2, \cdots x_i, \cdots x_{(n-1)}$. Let l be any integer less than n. Form the sum of the absolute values of $x_i - x_{(i-l)}$, designated by $\sum |x_i - x_{(i-l)}|$. Define $A = \sum_{i=l}^{n-1} \frac{|x_i - x_{(i-l)}|}{n - l}$, l takes the values of the various trial periods and is called the *lag*. A, therefore, is the mean error between prediction that data will be repeated after a lag of l and the fulfillment of the prediction. Such an index has a meaning that is immediately of use to a meteorologist or other investigator. Coefficients such as the Schuster and the correlation coefficient, although valuable statistically, are of less immediate interest.

The standard deviation of these errors of prediction follows at once from standard formulae under assumption of normal distribution.

$$\sigma = 1.25\,A$$

The distribution of σ, as computed from the absolute values of data, has been studied by Helmert and by Fisher. Davies and E. S. Pearson have compared the various methods of estimating σ. For the large number, $(n - l)$, pairs of data used for a periodogram point, this method becomes almost as precise as the usual one which would square the values of $(x_i - x_{i-l})$. For $(n - l)$ as small as 50, the standard deviation of the standard deviation by this method is only seven percent larger than by the other one. Fisher has shown that

$$\sigma_\sigma \to \frac{\sigma}{\sqrt{n-l}} \sqrt{\frac{\pi - 2}{2}} \qquad \text{as } (n - l) \to \infty$$

This may be written as

$$\sigma_\sigma \to \frac{1.068\,\sigma}{\sqrt{2(n - l)}}$$

The distribution approaches normal rapidly and for all values of $(n - l)$ that would be used in periodogram calculation certainly may be considered as normal. It will be very seldom that a value of $(n - l)$ much smaller than 200 will be used.

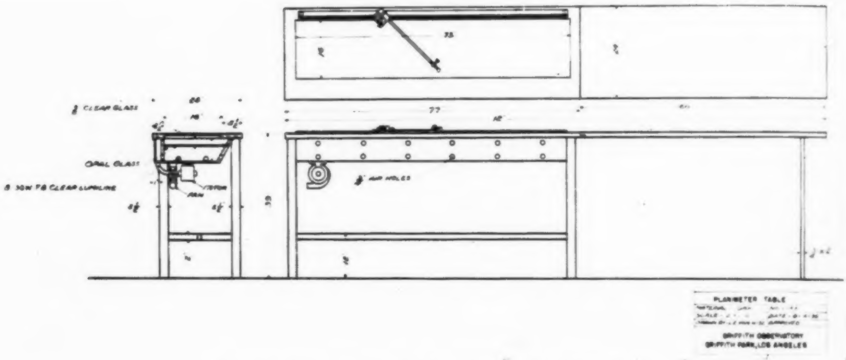
The data may be printed on two strips of adding machine tape held together by clips so as to match data separated by a lag l. In arranging them for investigation, it usually is most convenient to make all numbers positive. The computer subtracts mentally and puts the difference into an adding machine, which gives him A almost immediately.

For some computers, and especially where the numbers are large, another method of obtaining A may save time or lead to less numerical mistakes. The computer will form the sum of all his data. He will, as for the other form of computation, put these on two pieces of adding machine tape that he lays side by side. However, instead of putting the difference of the pairs into the machine, he will, in each case, put in the smaller datum of the pair. Then,

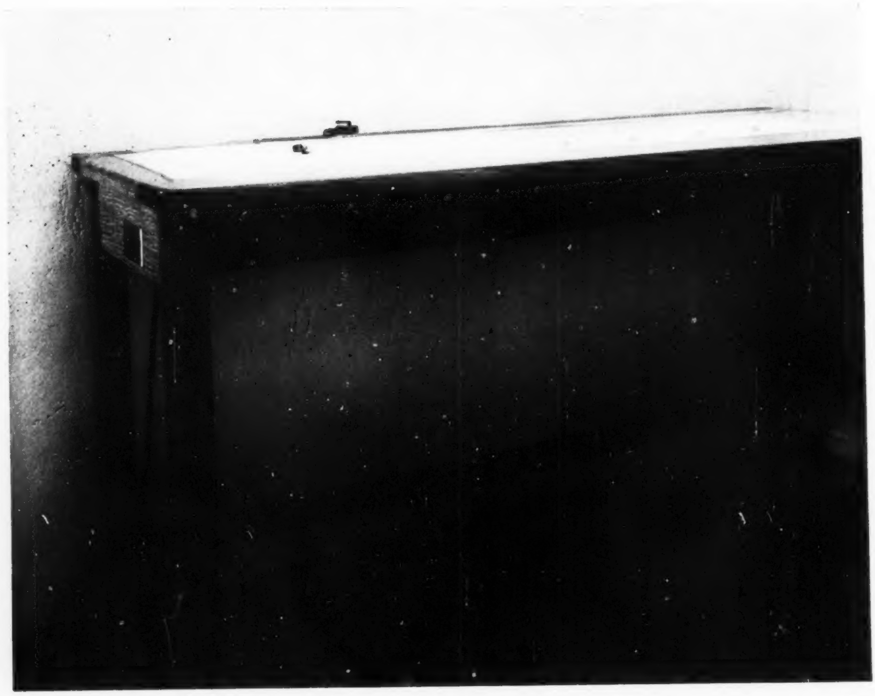
$$(n - l)A_l = 2 \sum \text{all data} - [\sum \text{1st } (n - l) + \sum \text{last } (n - l) \text{ data}] - 2 \sum \text{smaller}$$

The derivation of this equation is obvious. In computing by this method the subtotaler on the machine can be used to make the strip of sums of the first $(n - l)$ data and of the last $(n - l)$ for all values of l. The first term on the right hand side is a constant, the last is twice the sum of the smaller numbers chosen in the pairs. I have computed by both methods, and where the numbers are small, I prefer the former. Where they are large, I prefer the latter. However, when one must use comparatively untrained computers, he will find less mistakes made if the computer does not make the subtractions.

The calculation of A is much shorter than that for the indices even of the correlation and variance periodograms. It may, however, be shortened even more by a mechanical arrangement. $(n - l)A_l$ is the area between two histograms of the data matched after a lag l. These may be carefully graphed on a large scale and two such graphs superposed over a table with a translucent illuminated top. On the edge of this table is the track to guide a rolling planimeter. A, as computed by this means, is accurate to approximately one-half of one percent of its value, a much more exact value than is needed. The details of such a device as constructed for the Griffith Observatory are shown by the accompanying photograph and diagram. The dual saving of time by the method and by its mechanical application have resulted in the adoption of a much more ambitious program of meteorological research than previously was contemplated.



SCALE DIAGRAM OF PLANIMETER DEVICE



PLANIMETER DEVICE FOR MECHANICAL CALCULATION

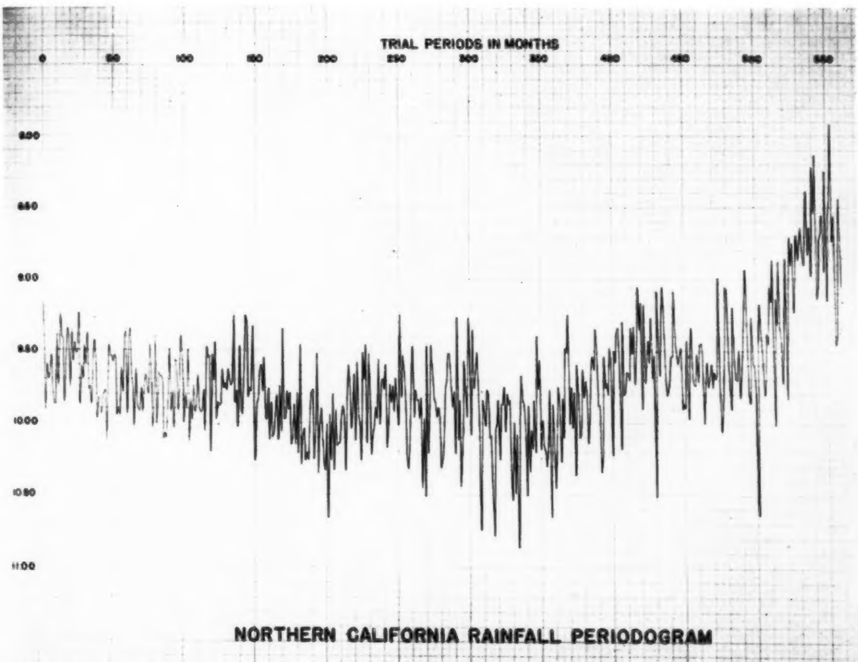
The form taken by the periodogram is important. Consider the simplest case, data which follow a sine curve.

$$y_i = a \cos\left(\frac{2\pi i - c}{p}\right)$$

$$y_i - y_{i-l} = 2a \sin \frac{\pi l}{p}\left[\sin \frac{2\pi(\frac{1}{2}l - i) + c}{p}\right]$$

The term in brackets takes values distributed around the circle and the part outside is a constant for any one lag. The bracket term sums approximately to $\frac{2(n-l)}{\pi}$, since we consider all terms as of one sign only.

$$\therefore A_l = \left|\frac{4a}{\pi} \sin \frac{\pi l}{p}\right|$$



NORTHERN CALIFORNIA RAINFALL PERIODOGRAM

If the absolute values were not considered in the expression for A_l, the periodogram would be a sine curve of period $2p$. The lack of sign gives a cusp curve with the cusp at lags p, $2p$, etc. Such a form is advantageous in that the periodogram gives sharp peaks at multiples of the periods which may exist.

The effect of the periodogram in exaggerating the principal terms at the expense of the smaller ones may be obtained most easily by equating σ as obtained by the linear and the quadratic formulae.

The data may be written as the sum of cosine terms

$$y_i = a\cos\left(\frac{2\pi i - \varphi_a}{p_a}\right) + b\cos\left(\frac{2\pi i - \varphi_b}{p_b}\right) + \cdots + c_i$$

$$y_i - y_{i-l} = 2a\sin\frac{\pi l}{p_a}\left[\sin\frac{2\pi(\frac{1}{2}l - i) + \varphi_a}{p_a}\right] + \cdots + (c_i - c_{i-l})$$

$$\sum (y_i - y_{i-l})^2 = 2(n-l)a^2\sin^2\frac{\pi l}{p_a} + 2(n-l)b^2\sin^2\frac{\pi l}{p_b} + \cdots + (n-l)\sqrt{2}\,\sigma_c^2$$

The sine terms contribute to A_l^2 in proportion to the squares of their amplitudes. On account of the $\sin^2\frac{\pi l}{p_j}$ factor, they contribute very little to values of A_l for which $\frac{\pi l}{p_j}$ is not very closely an even multiple of π.

This method has been applied to rainfall data of the Pacific Coast and has proved as satisfactory in practice as would be expected from the simplicity of the theory. The periodogram of rainfall stations along the northern third of the California coast is shown here, exhibiting perhaps the most definite single piece of evidence ever found for rainfall cycles. Outstanding is a cycle of about 45 years with its fourth harmonic as the secondary feature. The writer expects to publish the results of that work in the Monthly Weather Review.